# Gain-of-function of mutated C-CBL tumor suppressor in myeloid neoplasms( Dissertation_全文 )

AUTHOR(S):

Sanada, Masashi

*Nature* 460(7257):904-8. 主論文

# Gain-of-function of mutated *C-CBL* tumor suppressor in myeloid neoplasms

Masashi Sanada[1,2]*, Takahiro Suzuki[3]*, Lee-Yung Shih[4]*, Makoto Otsu[5], Satoshi Yamazaki[5], Motohiro Kato[1,6], Azusa Tamura[1], Hiroaki Honda[7], Mamiko Sakata-Yanagimoto[8], Keiki Kumano[9], Hideaki Oda[10], Tetsuya Yamagata[11], Junko Takita[1,6,9], Noriko Gotoh[12], Kumi Nakazaki[1,13], Norihiko Kawamata[14], Masafumi Onodera, Masaharu Nobuyoshi[3], Yasuhide Hayashi[15], Hiroshi Harada[16], Mineo Kurokawa[9,13], Shigeru Chiba[8], Hiraku Mori[16], Keiya Ozawa[3], Mitsuhiro Omine[16], Hisamaru Hirai[9,13], Hiromitsu Nakauchi[5], H. Phillip Koeffler[14], Seishi Ogawa[1,2]

[1]Cancer Genomics Project, Department of [6]Pediatrics, [9]Cell Therapy and Transplantation Medicine and [13]Hematology and Oncology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan
[2]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi-shi, Saitama 332-0012, Japan
[3]Division of Hematology, Department of Medicine, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke-shi, Tochigi 329-0498, Japan
[4]Division of Hematology-Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University, 199 Tung Hwa North Rd, Taipei 105, Taiwan
[5]Laboratory of Stem Cell Therapy, Center for Experimental Medicine, [12]Division of Systems Biomedical Technology, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[7]Department of Developmental Biology, Research Institute of Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan
[8]Department of Clinical and Experimental Hematology, Institute of Clinical Medicine, University of Tsukuba, 1-1-1 Tennodai, Tukuba-shi, Ibaragi, 305-8571, Japan
[10]Department of Pathology, Tokyo Women's Medical University, 8-1 Kawada-cho, Shinjuku-ku, Tokyo 162-8666, Japan
[11]Department of Hematology, Dokkyo University School of Medicine, 800 Kitabayashi, Mibu, Tochigi 321-0293, Japan
[14]Hematology/Oncology, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
[15]Gunma Children's Medical Center, 779 Shimohakoda, Hokkitsu-machi, Shibukawa-shi, Gunma 377-8577, Japan
[16]Division of Hematology, Internal Medicine, Showa University Fujigaoka Hospital, 1-30 Fujigaoka, Aoba-ku, Yokohama, Kanagawa 227-8501, Japan

* These authors contributed equally to this work.

Correspondence to:
Seishi Ogawa, M.D., Ph.D. (sogawa-tky@umin.ac.jp) and Lee-Yung Shih, M.D. (sly7012@adm.cgmh.org.tw)

This paper was published in Nature 2009 Aug 13;460(7257):904-8.
doi: 10.1038/nature08240.
**http://www.nature.com/nature/journal/v460/n7257/full/nature08240.html**

**Summary**

**Acquired uniparental disomy (aUPD) is a common feature of cancer genomes, leading to loss of heterozygosity (LOH). aUPD is associated not only with loss-of-function mutations of tumor suppressor genes[1] but also with gain-of-function mutations of proto-oncogenes[2]. Here we show unique gain-of-function mutations of *C-CBL* (also known as *CBL*) tumor suppressor that are tightly associated with aUPD of the 11q arm in myeloid neoplasms showing myeloproliferative features. The *C-CBL* proto-oncogene, a cellular homolog of *v-Cbl*, encodes an E3 ubiquitin ligase and negatively regulates signal transduction of tyrosine kinases[3-6]. Homozygous *C-CBL* mutations were found in most 11q-aUPD-positive myeloid malignancies. Although the *C-CBL* mutations were oncogenic in NIH3T3 cells, c-Cbl was shown to functionally and genetically act as a tumor suppressor. C-CBL mutants did not have E3 ubiquitin ligase activity, but inhibited that of wild-type C-CBL and CBL-B (also known as CBLB), leading to prolonged activation of tyrosine kinases after cytokine stimulation. *c-Cbl*$^{-/-}$ hematopoietic stem/progenitor cells (HSPCs) showed enhanced sensitivity to a variety of cytokines compared to *c-Cbl*$^{+/+}$ HSPCs, but transduction of *C-CBL* mutants into *c-Cbl*$^{-/-}$ HSPCs further augmented their sensitivities to a broader spectrum of cytokines, including the stem-cell factor (SCF), thrombopoietin (TPO), IL3, and FLT3 ligand (FLT3LG), indicating the presence of gain-of-function that could not be attributed to a simple loss-of-function. The gain-of-function effects of C-CBL mutants on cytokine sensitivity of HSPCs largely disappeared in a *c-Cbl*$^{+/+}$ background or by co-transduction of wild-type *C-CBL*, which suggests the pathogenic importance of loss of wild-type *C-CBL* alleles found in most cases of *C-CBL*-mutated myeloid neoplasms. Our findings provide a novel insight into a role of gain-of-function mutations of a tumor suppressor associated with aUPD in the pathogenesis of some myeloid cancer subsets.**

Myelodysplastic syndromes (MDS) are heterogeneous groups of blood cancers originated from hematopoietic precursors. They are characterized by deregulated hematopoiesis showing a high propensity to acute myeloid leukemia (AML)[7]. Some MDS cases have overlapping clinic-pathological features with myeloproliferative disorders and are now classified into myelodysplasia/myeloproliferative neoplasms (MDS/MPN) in the World Health Organization (WHO) classification[8]. To obtain a comprehensive profile of allelic imbalances in these myeloid neoplasms, we performed allele-specific copy number analyses of bone marrow samples obtained from 222 patients with MDS, MDS/MPN, or other related myeloid neoplasms (Supplementary Tables 1 and 2) using high-density single nucleotide polymorphism (SNP) arrays combined with CNAG/AsCNAR software[9,10].

Genomic profiles of MDS and MDS/MPN revealed characteristic unbalanced genetic changes, as reported in previous cytogenetic studies[11] (Supplementary Fig. 1a), however, they were detected more sensitively by SNP array analyses (Supplementary Table 3). aUPD was detected in 70 samples (31.5%) on the basis of the allele-specific copy number analyses, which substantially exceed the detection rate obtained using a SNP call-based detection algorithm (20.7%) (Supplementary Figs. 2 and 4, and Supplementary Table 4). Long stretches of homozygous SNP calls caused by shared identical-by-descent (IBD) alleles in parents were empirically predicted and excluded (Supplementary Fig. 3). aUPDs were more common in MDS/MPN than MDS. They preferentially affected several chromosomal arms (1p, 1q, 4q, 7q, 11p, 11q, 14q, 17q, and 21q) in distinct subsets of patients, and frequently associated with mutated oncogenes and tumor suppressor genes (Supplementary Figs. 1b and 5). Among these, the most common aUPDs were those involving 11q (n = 17), which defined a unique subset of myeloid neoplasms that were clinically characterized by frequent diagnosis of chronic myelomonocytic leukemia (CMML) with normal karyotypes (13 cases) (Fig. 1a

and Supplementary Table 6). We identified a minimum overlapping aUPD segment of approximately 1.4 megabases (Mb) in 11q, which contained a mutated *C-CBL* proto-oncogene (Fig. 1b).

*C-CBL* is the cellular homolog of the *v-Cbl* transforming gene of the Cas NS-1 murine leukemia virus[5,12]. It was recently found to be mutated in human AML cases[13-15]. Together with its close homolog, CBL-B, C-CBL is thought to be engaged in the negative modulation of tyrosine kinase signaling, primarily through their E3 ubiquitin ligase activity that is responsible for the downregulation of activated tyrosine kinases[3-5]. By sequencing all *C-CBL* exons in all 222 samples, we found *C-CBL* mutations in 15 of the 17 cases with 11q-aUPD, whereas only 3 out of 205 cases without 11q-aUPD had *C-CBL* mutations, showing a strong association of *C-CBL* mutations with 11q-aUPD ($p = 1.46 \times 10^{-18}$) (Supplementary Fig. 6 and Supplementary Tables 6 and 7), as also indicated in a recent report[16]. Thus, *C-CBL* was thought to be the major, if not only, target of 11q-aUPD in myeloid neoplasms. Two different *C-CBL* mutations coexisted in three cases (Supplementary Fig. 6b). Somatic origins of the mutations were confirmed in three evaluable cases (Supplementary Fig. 6c).

In most cases, *C-CBL* mutations were missense, involving the evolutionarily conserved amino acids within the linker-RING finger domain that is central to the E3 ubiquitin ligase activity[17] (Fig. 1c). Another case having a predominant Cys384Tyr mutation also contained a nonsense mutation (R343X) in a minor subclone, which resulted in a v-Cbl-like truncated protein (Supplementary Fig. 6b). In the remaining two cases, mutations lead to amino acid deletions (Δ368-371 and Δ368-382) involving the highly conserved α-helix (αL) of the linker domain and the first loop of the RING finger. According to the published crystal structure of C-CBL[17], most of the mutated or deleted amino acids were positioned on the interface for the binding to E2 enzymes (Fig. 1d), making contact with either the tyrosine kinase binding domain (Tyr368 and Tyr371) or E2 ubiquitin-conjugating enzymes (Ile383, Cys404, and Phe418). Especially, all

seven linker domain mutations selectively involved just three amino acids (Glu367, Tyr368, and Tyr371) within the conserved αL helix (Fig. 1d). Mutations were clearly homozygous in 9 cases, and the apparently heterozygous chromatograms in the other 6 cases could also be compatible with homozygous mutations affecting the aUPD-positive tumor clones, given the presence of substantial normal cell components within these samples (Supplementary Fig. 6a). Mutations in the remaining 3 cases were considered to be heterozygous. About half of the *C-CBL*-mutated cases carried coexisting mutations of *RUNX1* (4 cases), *TP53* (1 case), *FLT3* internal tandem duplication (1 case) or *JAK2* (3 cases). No coexisting *NRAS* and *KRAS* mutations were prevalent among CMML (15.1 %) but occurred within discrete clusters from *C-CBL*-mutated cases (Supplementary Tables 2 and 6 and Supplementary Fig. 5). The mutation status of *C-CBL* did not substantially affect the clinical outcome (Supplementary Fig. 7).

All tested *C-CBL* mutants induced clear oncogenic phenotypes in NIH3T3 fibroblasts, as demonstrated by enhanced colony formation in soft agar and tumor generation in nude mice (Supplementary Fig. 8). Transformed NIH3T3 cells showed PI3 kinase-dependent activation of Akt and the transformed phenotype was reverted by treatment with increasing concentrations of a PI3 kinase inhibitor LY294002 (Supplementary Fig. 9). When introduced into $Lin^{-}Sca1^{+}c\text{-}Kit^{+}$ (LSK) HSPCs, C-CBL mutants ($C\text{-}CBL^{Glu367Pro}$ and $C\text{-}CBL^{Tyr371Ser}$), as well as a mouse lymphoma-derived oncogenic mutant (70Z mutant, $C\text{-}CBL^{70Z}$), significantly promoted replating capacity of these progenitors (Fig. 2a). Because C-CBL negatively modulates tyrosine kinase signaling, and all C-CBL mutations, including those previously reported[13-16] affected the critical domains for its enzymatic activity involved in this modulation, C-CBL was postulated to have a tumor suppressor function; loss-of-function could be a mechanism for the oncogenicity of these C-CBL mutants[3,5]. To assess this possibility and to clarify further the role of *C-CBL* mutations in the pathogenesis of myeloid neoplasms, we generated $c\text{-}Cbl^{-/-}$ mice and examined their hematologic phenotypes (Supplementary

Fig. 10).

In accordance with previous reports[18-20], *c-Cbl*$^{-/-}$ mice exhibited splenomegaly and an augmented hematopoietic progenitor pool as was evident from the increased colony formation of bone marrow cells in methylcellulose culture and higher numbers of LSK and CD34-negative LSK cells in bone marrow and/or spleen compared to their wild-type littermates (Figs. 2b-d and Supplementary Fig. 11). Furthermore, when introduced into a *BCR/ABL* transgenic background[21], the *c-Cbl*$^{-/-}$ allele accelerated blastic crisis depending on the allele dosage (Figs. 2e,f). These observations supported the notion that wild type *C-CBL* has tumor suppressor functions, whereas "mutant" *C-CBL* acts as an oncogene; *C-CBL* can therefore be both a proto-oncogene and a tumor suppressor gene.

Mouse LSK HSPCs expressed two Cbl family member proteins: wild-type c-Cbl and Cbl-b (Supplementary Fig. 12)[22]. When transduced into NIH3T3 cells stably expressing human epidermal growth factor receptor (EGFR), both Cbl proteins enhanced ubiquitination of EGFR after EGF-stimulation, which was suppressed by coexpression of the C-CBL mutants (Figs. 3a,b). In hematopoietic cells, over-expression of wild-type C-CBL enhanced ligand-induced ubiquitination of a variety of tyrosine kinases, including c-KIT, FLT3, and JAK2. In contrast, C-CBL mutants not only showed compromised enzymatic activity, but also inhibited the ubiquitinating activities in these hematopoietic cells (Fig. 3c), leading to prolonged tyrosine kinase activation after ligand stimulation (Fig. 3d).

Because tyrosine kinase signaling is central to cytokine responses in hematopoietic cells and its deregulation is a common feature of myeloproliferative disorders[23], we next examined the effects of *C-CBL* mutations (*C-CBL*$^{Glu367Pro}$ and *C-CBL*$^{Tyr371Ser}$) and loss of wild-type *C-CBL* alleles on the responses of LSK HSPCs to various cytokines. In serum-free conditions, *c-Cbl*$^{-/-}$ LSK cells showed a modestly enhanced proliferative response to a variety of cytokines, including SCF, IL3, and TPO, compared to *c-Cbl*$^{+/+}$

cells (mock columns in Fig. 4a). However, the enhanced response in *c-Cbl*$^{-/-}$ cells was markedly augmented and extended to a broader spectrum of cytokines, including FLT3 ligand by the transduction of *C-CBL* mutants. Of note, the effect of *C-CBL* mutant transduction was not remarkable in *c-Cbl*$^{+/+}$ LSK cells except for the response to SCF, which was clearly enhanced by *C-CBL* mutants even with a *c-Cbl*$^{+/+}$ background (Fig. 4a and Supplementary Fig. 13). To clarify further the effect of wild-type C-CBL on C-CBL mutants, both wild-type *C-CBL* and *C-CBL* mutants were co-transduced into *c-Cbl*$^{-/-}$ LSK cells, and their effects on the response to SCF were examined. As shown in Fig. 4b, the hyperproliferative response induced by *C-CBL* mutants was almost completely abolished by the co-transduction of wild-type *C-CBL*, suggesting the pathogenic importance of loss of wild-type *C-CBL* alleles found in most *C-CBL*-mutated cases. LSK cells transduced with *C-CBL* mutants also showed enhanced activation of the STAT5 and Akt pathways on cytokine stimulation (SCF and TPO), which was more pronouced in *c-Cbl*$^{-/-}$ than *c-Cbl*$^{+/+}$ LSK cells (Fig. 4c and Supplementary Fig. 14).

The modest enhancement of sensitivity to cytokines found in *c-Cbl*$^{-/-}$ LSK cells was a consequence of loss of C-CBL functions. In contrast, the hypersensitive response of mutant-transduced *c-Cbl*$^{-/-}$ LSK cells to a broad spectrum of cytokines represents gain-of-function of the mutants that could not be ascribed to simple loss of C-CBL functions, which was also predicted from the strong association of *C-CBL* mutations with 11q-aUPD by analogy to the gain-of-function *JAK2* mutations associated with 9p-aUPD in polycythemia vera[2]. The gain-of-function of C-CBL mutants becomes more evident under a *c-Cbl*$^{-/-}$ background. The hypersensitive response to cytokines induced by mutant C-CBL under the *c-Cbl*$^{-/-}$ background was largely offset by the presence of the wild-type *C-CBL* allele or by the transduction of wild-type *C-CBL* gene, suggesting that the gain-of-function could be closely related to loss of C-CBL-like functions, probably by inhibition of Cbl-b. Supporting this view is a previous report that

*c-Cbl/Cbl-b* double knockout T-cells showed more profound impairments in the down-regulation of the T-cell receptor (TCR), more sustained TCR signaling, and more vigorous proliferation, than *c-Cbl* or *Cbl-b* single knockout T-cells after anti-CD3 (also known as CD3e) stimulation[24]. This is analogous to the gain-of-function found in some TP53 mutants, which has been explained by functional inhibition of two TP53 homologs, TP73 and TP63[25,26]. Of note, TP53 was also originally isolated as an oncogene through its mutated forms[27]. The Cbl-b inhibition-based gain-of-function model could be tested directly by comparing the behavior of *c-Cbl/Cbl-b* double knockout LSK cells with that of LSK cell carrying homozygously knocked-in mutant *C-CBL* alleles. On the other hand, there remains a possibility that the gain-of-function could be mediated by a mechanism other than the simple inhibition of the homolog, because C-CBL mutants retained several motifs that interacted with numerous signal-transducing molecules. Furthermore, considering the ubiquitous expression of CBL proteins, it would be of interest to explore the possible involvement of mutations in all CBL family members in other human cancers.

## METHODS SUMMARY

Genomic DNAs from 222 bone marrow samples with myeloid neoplasms were analyzed using GeneChip SNP-genotyping microarrays (Affymetrix GeneChip®) as described[28]. Allelic imbalances were detected from the allele-specific copy numbers calculated using CNAG/AsCNAR software (http://www.genome.umin.jp)[9,10]. *C-CBL* mutations were examined by sequencing PCR-amplified genomic DNAs. For functional assays, HA- or FLAG-tagged cDNAs of wild-type and mutant *C-CBL* were generated by *in vitro* mutagenesis, constructed into a MSCV-based retroviral vector, pGCDNsamIRESGFP or pGCDNsamIRESKO, and used for retrovirus-mediated gene transfer. For the evaluation of oncogenicity of C-CBL mutants, NIH3T3 cells were transfected with various *C-CBL* constructs and subjected to colony assays in soft agar and tumor formation assays in nude mice. *c-Cbl*-deficient mice were generated according to the conventional strategy of gene-targeting and crossed with *bcr/abl* transgenic mice to evaluate the effect of *c-Cbl*$^{-}$ allele on the acceleration of blastic crisis. LSK cells sorted from *c-Cbl*$^{+/+}$ and *c-Cbl*$^{-/-}$ mice were transduced with various *C-CBL* constructs and their responses to cytokines were evaluated by cell proliferation assays followed by immunoblot analyses of c-Kit, FLT3, and JAK2, and their downstream signaling molecules. The effects of expression of C-CBL mutants on ubiquitination of EGFR, c-KIT, FLT3, and JAK2 were examined by transducing C-CBL mutants into relevant cells, followed by anti-ubiquitin blots of the immunoprecipitated kinases after ligand stimulation. The functional competition of C-CBL mutants with wild-type C-CBL was assessed by cell proliferation assays of LSK cells cotransduced with both c-DNAs. This study was approved by the ethics boards of the University of Tokyo, Chang Gung Memorial Hospital, and Showa University. Antibodies and primers used in this study are listed in Supplementary Tables 8 and 9.

## Full Methods

### Genome-wide analysis of allelic imbalances in primary myeloid neoplasms

BM specimens were obtained from 222 patients diagnosed as having myeloid neoplasms according to the WHO classification (Supplementary Table 1 and 2), from which high-molecular-weight genomic DNA was extracted and subjected to microarray analysis using Affymetrix GeneChip® 50K *Xba*I, *Hin*dIII, or 250K *Nsp*I, according to the manufacturer's protocol. Genome-wide detection of allelic imbalances was performed using CNAG/AsCNAR software (http://www.genome.umin.jp)[9,10].

### Mutation analysis

Mutation analysis was performed by direct sequencing of PCR-amplified coding exons of the relevant genes using an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems). The target genes and exons as well as the PCR primers are listed in Supplementary Table 8. Tandem duplication of *FLT3* gene was examined by genomic PCR and sequencing.

### Preparation of high-titer VSV-G-pseudotyped retroviral particles

The HA-tagged human *C-CBL* cDNA was a kind gift from Dr. W.Y. Langdon. In total, 9 mutant c-DNAs of *C-CBL*, including 8 from patients' specimens and a 70Z mutant corresponding to a mutant isolated from mouse lymphoma[29], were generated based on this construct using a QuickChange® Site-Directed Mutagenesis Kit® (Stratagene) and constructed into retrovirus vectors pGCDNsamIRESGFP and pGCDNsamIRESKO[30,31]. Vector plasmids were cotransfected with a VSVG cDNA into 293GP cells (kindly provided by Dr. R.C. Mulligan) to obtain retrovirus-containing supernatant, which was then transduced into 293GPG cells to establish stable cell lines capable of producing VSV-G pseudotyped retroviral particles upon induction[32,33]. The average titer of retrovirus stocks prepared from these cell lines routinely exceeded ~1–10 × $10^7$ ifu/mL, as estimated using Jurkat cells.

### Assays for anchorage independent growth and tumorigenicity in nude mice

NIH3T3 cells (the Japan Cell Resource Bank) were stably transduced with wild-type and mutant *C-CBL-B*y retrovirus-mediated gene transfer. For colony formation assays, $1.0 \times 10^3$ stable cells for each construct were inoculated in 0.33% top agar, and the numbers of colonies larger than 1 mm in diameter were counted 3 weeks after inoculation (N = 8). Experiments were repeated four times. For tumor formation in nude mice, $1.0 \times 10^7$ stable cells were inoculated subcutaneously at 2 sites per mouse. Cells were inoculated at 6 sites in 3 mice for each construct.

**Purification of KSL hematopoietic stem/progenitor cells (HSPCs)**

$Lin^-Sca1^+c\text{-}Kit^+$ HSPCs were purified from BM and spleen as described[14,34]. Multicolor flow cytometry analysis and cell-sorting were performed using a MoFlo Cell Sorter (Beckman Coulter). Purity of sorted cell fractions consistently exceeded 98%.

**Replating assays of BM progenitor cells**

BM LSK cells were infected with IRES/GFP-containing retrovirus carrying, mock, *C-CBL*$^{WT}$, and 3 *C-CBL* mutants (*C-CBL*$^{Q367P}$, *C-CBL*$^{Y371S}$, and *C-CBL*$^{C384G}$), as well as *C-CBL*$^{70Z}$ on RetroNectin™-coated dishes. After 48 h infection in culture in StemSpan® supplemented with SCF (50 ng/mL, Peprotech), TPO (20 ng/mL), and FLT3 ligand (20 ng/mL), $1.0 \times 10^2$ GFP-positive cells were inoculated in MethoCult® M3231 supplemented with thrombopoietin (20 ng/mL), IL-3 (10 ng/mL), IL-6 (10 ng/mL), FLT3 ligand (10 ng/mL), and stem-cell factor (50 ng/mL) for colony formation. Seven days after each inoculation, colony-forming cells were harvested, from which $1.0 \times 10^3$ cells were repeatedly subjected to replating until no colonies were produced. Experiments were repeated four times for each *C-CBL* construct.

**Generation of *c-Cbl*$^{-/-}$ mice and evaluation of their tumor-prone phenotype**

*c-Cbl*$^{-/-}$ mice were generated according to the conventional method of gene targeting (Supplementary Fig. 10). *c-Cbl*$^{+/+}$, *c-Cbl*$^{+/-}$, and *c-Cbl*$^{-/-}$ mice were crossed with *bcr/abl* transgenic mice, and their survival and the development of blastic crises were monitored.

**Evaluation of hematopoietic pool size in *c-Cbl*$^{-/-}$ mice**

Lin$^{-}$Sca1$^{+}$c-Kit$^{+}$ (LSK) and CD34$^{-}$LSK cells were sorted from BM cells or spleen of *c-Cbl*$^{-/-}$ mice, and their numbers were compared with those in *c-Cbl*$^{+/+}$. Approximately 5 × $10^3$ of BM cells harvested from *c-Cbl*$^{+/+}$ and *c-Cbl*$^{-/-}$ mice, were inoculated into MethoCult® M3231 culture supplemented with TPO (20 ng/mL), IL-3 (10 ng/mL), IL-6 (10 ng/mL), erythropoietin (3U/mL), and stem-cell factor (50 ng/mL), and the numbers of colonies were counted 7 days after culturing.

***In vitro* cell proliferation assays**

Approximately 6 × $10^3$ LSK cells from *c-Cbl*$^{-/-}$ mice and their *c-Cbl*$^{+/+}$ littermates (8 w/o) were sorted into Retronectin$^{TM}$-coated 96-well U-bottom plates containing α-MEM supplemented with 1% fetal bovine serum (FBS), mouse SCF (50 ng/mL), and human TPO (100 ng/mL). After 24 h pre-incubation, retrovirus supernatant was added to each well at a multiplicity of infection (M.O.I.) of ~10, and the plates were incubated for another 24 h in the presence of protamine sulfate (10 μg/mL), followed by repeated infection and extended culture for 2 days in S-Clone SF-O3 medium (Sanko Junyaku) supplemented with 1% bovine serum albumin (BSA), 50 ng/mL SCF, and 50 ng/mL TPO. On day 4, fluorescent marker-positive cells were sorted for subsequent analyses. Cell survival and proliferation of LSK cells transduced with different *C-CBL* constructs was assessed in serum-free liquid culture in 96-well U-bottom plates in the presence of various cytokines. Each well received 50 fluorescent marker-positive LSK cells, and the cells were cultured in S-Clone supplemented with 1% BSA plus SCF, TPO, IL3, or FL, at the indicated concentrations. Cell numbers were counted either by analyzing well images or by flow cytometry using FlowCount beads (Beckman Coulter). After 6 h serum-starvation, 1 × $10^4$ LSK cells transduced with various *C-CBL* constructs were stimulated with SCF (10 ng/mL) and TPO (10 ng/mL) for 15 min. Whole-cell lysates were examined for activation of c-Kit, STAT5, and Akt by immunoblots using the indicated antibodies.

**Immunoblot analysis of physical interaction between mutant C-CBL and CBL-B**

Either FLAG-tagged CBL-B or FLAG-tagged C-CBL was cotransfected into NIH3T3 cells with each of three HA-tagged *C-CBL* mutants (*C-CBL*$^{Q367P}$, *C-CBL*$^{Y371S}$, and *C-CBL*$^{70Z}$). Total cell lysates of these NIH3T3 cells were immunoprecipitated with anti-FLAG antibody, followed by immunoblot analysis with anti-HA antibody.

**Detection of ubiquitination and phosphorylation of kinases**

After overnight serum starvation, NIH3T3 cells stably transduced with hEGFR and indicated HA-tagged *C-CBL* mutants and FLAG-tagged *C-CBL*$^{WT}$ were stimulated with hEGF (10 ng/mL) for 2 min. Cell lysates were immunoprecipitated with anti-hEGF antibody, followed by immunoblotting using anti-ubiquitin antibody (α-Ub). Constructs for C-CBL$^{WT}$ and mutant-Cbl were stably transduced into a mast-cell line, V3MC, FLT3-transduced 32D cells (32D/FLT3), and BaF3 cells transduced with hEpoR and JAK2 (BaF3/Epo/JAK2) using retrovirus-mediated gene transfer. After overnight serum starvation, the transduced cells were stimulated either with 10 ng/mL SCF (V3MC), 10 U/mL hEpo (BaF3/Epo/JAK2), or 10 ng/mL FL (32D/FLT3) for 1 min. The relevant kinases were immunoprecipitated with indicated antibodies, and their ubiquitination was detected by immunoblotting with α-Ub. Tyrosine phosphorylation of these kinases were examined by 4G10 blots of immunoprecipitated kinases at indicated time points. Antibodies used in this study are listed in Supplementary Table 9.

**Statistical Analysis**

Significance of prolonged replating capacity of mutant *C-CBL*-transduced LSK cells was tested by counting the total number of dishes that produced colonies, followed by Fisher's exact test. Survival curves of *c-Cbl*$^{+/+}$, *c-Cbl*$^{+/-}$, and *c-Cbl*$^{-/-}$ mice having the *bcr/abl* transgene were drawn following the Kaplan–Meier method. Overall survivals of C-CBL mutated and non-mutated CMML cases were analyzed according to the proportional hazard model using STATA software®. Statistical differences in survival were evaluated using a log rank test. Statistical differences in 2 × 2 contingency tables

were tested according to Fisher's exact method. Student t-tests were used to evaluate the significance of difference in spleen mass, numbers of hematopoietic progenitors, and colony forming cells between *c-Cbl*$^{+/+}$ and *c-Cbl*$^{-/-}$.

## References


## Acknowledgments

This work was supported by the Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, by a Grant-in-Aid from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology, grants from National Health Research Institute, and grants NHRI-EX96-9434SI and NIH- 2R01CA026038-30A1. We thank Drs. Wallace Y. Langdon for providing a human *C-CBL* cDNA. A mast-cell cell line expressing c-Kit, V3MC, was a generous gift from Dr. Michael F. Gurish.


## Author Contributions

M.S. and M.Kato performed microarray experiments and subsequent data analyses. M.S., T.S., T.Y., H. Honda, and H. Hirai generated and analyzed *c–Cbl* null mice. M.S., M. Otsu, S.Y., M.N., K.K., N.G., M.Onodera, M.S–Y., and H.N. conducted functional assays of C-CBL mutants. M.S., K.N., J.T., and A.T. performed mutation analysis. H.O. performed pathological analysis of *c–Cbl* null mice. L.S., N.K., H. Harada, M. Kurokawa, S.C., H.M., and M.Omine prepared MDS specimens. M.S., L.S., M. Otsu, Y.H., K.O., H.M., M. Omine, H.N., HP.K., and S.O. designed the overall study, and S.O. wrote the manuscript. All authors discussed the results and commented on the manuscript.

## AUTHOR INFORMATION

Full copy number data for the 222 samples are accessible from http://ncbi.nlm.nih.gov/geo/ with the accession number GSE15187. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.O. (sogawa–tky@umin.ac.jp) and LY.S. (sly7012@adm.cgmh.org.tw).

Figure 1

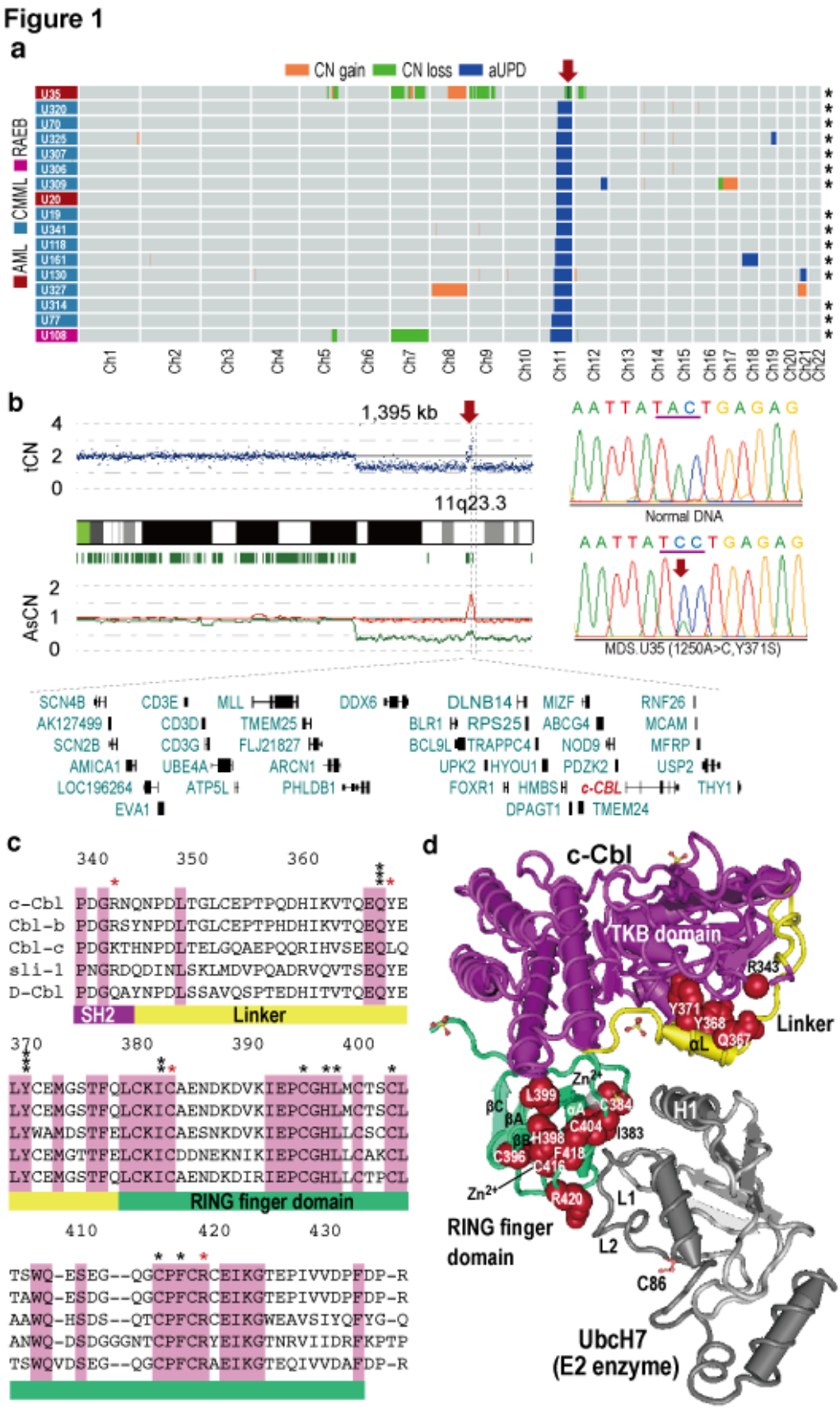

a
CN gain
CN loss
aUPD
RAEB
CMML
AML
U35
U320
U70
U326
U307
U306
U309
U20
U19
U341
U118
U161
U130
U327
U314
U77
U108
Ch1
Ch2
Ch3
Ch4
Ch5
Ch6
Ch7
Ch8
Ch9
Ch10
Ch11
Ch12
Ch13
Ch14
Ch15
Ch16
Ch17
Ch18
Ch19
Ch20
Ch21
Ch22
b
1,395 kb
tCN
11q23.3
AsCN
AATTATACTGAGAG
Normal DNA
AATTATCCTGAGAG
MDS.U35 (1250A>C,Y371S)
SCN4B
CD3E
MLL
DDX6
DLNB14
MIZF
RNF26
AK127499
CD3D
TMEM25
BLR1
RPS25
ABCG4
MCAM
SCN2B
CD3G
FLJ21827
BCL9L
TRAPPC4
NOD9
MFRP
AMICA1
UBE4A
ARCN1
UPK2
HYOU1
PDZK2
USP2
LOC196264
ATP5L
PHLDB1
FOXR1
HMBS
c-CBL
THY1
EVA1
DPAGT1
TMEM24
c
340
350
360
c-Cbl PDGRNQNPDLTGLCEPTPQDHIKVTQEQYE
Cbl-b PDGRSYNPDLTGLCEPTPHDHIKVTQEQYE
Cbl-c PDGKTHNPDLTELGQAEPQQRIHVSEEQLQ
sli-1 PNGRDQDINLSKLMDVPQADRVQVTSEQYE
D-Cbl PDGQAYNPDLSSAVQSPTEDHITVTQEQYE
SH2
Linker
370
380
390
400
LYCEMGSTFQLCKICAENDKDVKIEPCGHLMCTSCL
LYCEMGSTFQLCKICAENDKDVKIEPCGHLMCTSCL
LYWAMDSTFELCKICAESNKDVKIEPCGHLLCSCCL
LYCEMGTTFELCKICDDNEKNIKIEPCGHLLCAKCL
LYCEMGSTFQLCKICAENDKDIRIEPCGHLLCTPCL
RING finger domain
410
420
430
TSWQ-ESEG--QGCPFCRCEIKGTEPIVVDPFDP-R
TAWQ-ESDG--QGCPFCRCEIKGTEPIVVDPFDP-R
AAWQ-HSDS--QTCPFCRCEIKGWEAVSIYQFYG-Q
ANWQ-DSDGGGNTCPFCRYEIKGTNRVIIDRFKPTP
TSWQVDSEG--QGCPFCRAEIKGTEQIVVDAFDP-R
d
c-Cbl
TKB domain
R343
Y371
Y368
αL
Q367
Linker
Zn2+
βC
βA
L399
αA
C384
H1
C404
βB
H398
I383
C396
F418
C416
Zn2+
R420
L1
RING finger
domain
L2
C86
UbcH7
(E2 enzyme)

Figure 2

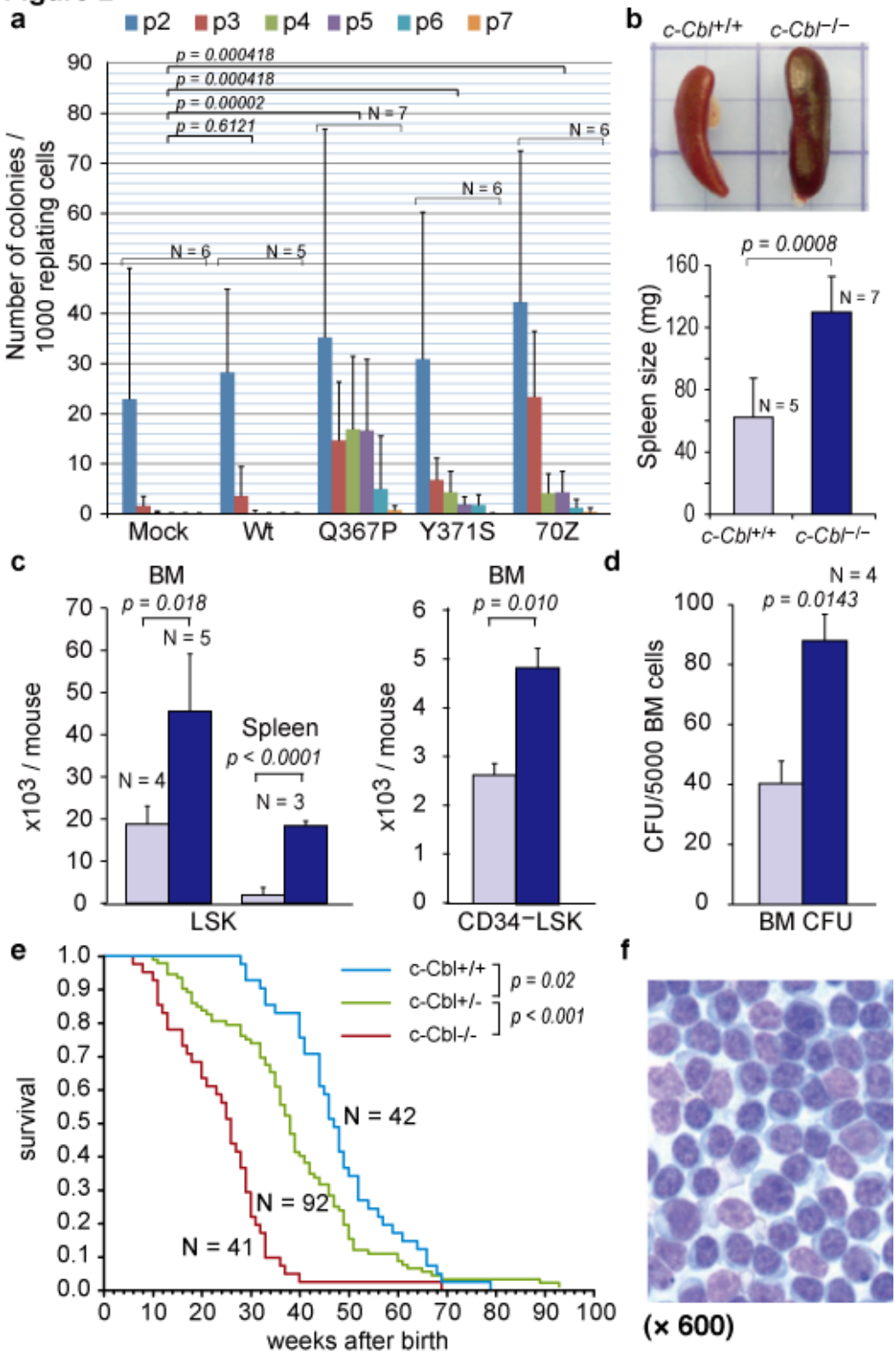

a
p2
p3
p4
p5
p6
p7
Number of colonies / 1000 replating cells
p = 0.000418
p = 0.000418
p = 0.00002
p = 0.6121
N = 7
N = 6
N = 6
N = 6
N = 5
Mock
Wt
Q367P
Y371S
70Z
b
c-Cbl+/+
c-Cbl−/−
p = 0.0008
Spleen size (mg)
N = 5
N = 7
c-Cbl+/+
c-Cbl−/−
c
BM
p = 0.018
N = 5
N = 4
Spleen
p < 0.0001
N = 3
x10³ / mouse
LSK
BM
p = 0.010
x10³ / mouse
CD34−LSK
d
N = 4
p = 0.0143
CFU/5000 BM cells
BM CFU
e
c-Cbl+/+
c-Cbl+/-
c-Cbl-/-
p = 0.02
p < 0.001
N = 42
N = 92
N = 41
survival
weeks after birth
f
(× 600)

**Figure 3**

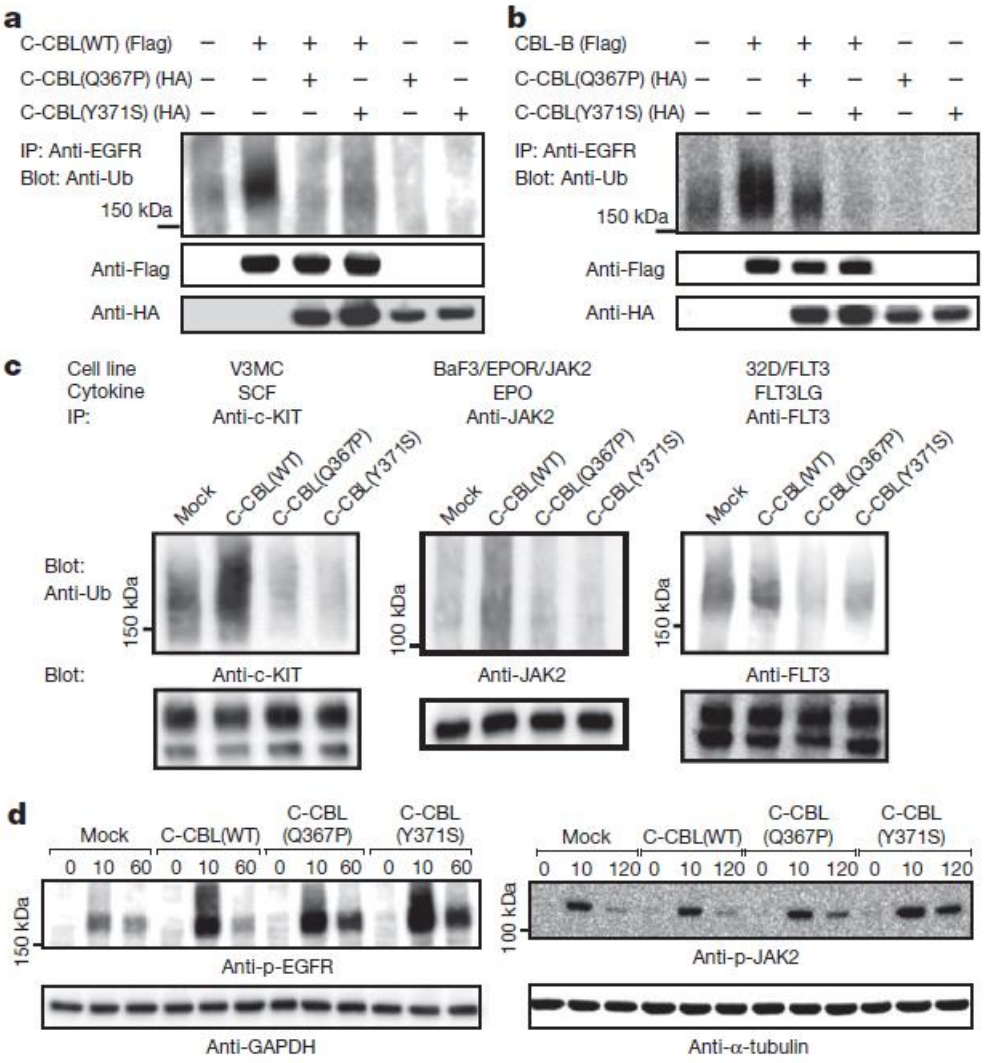

a
C-CBL(WT) (Flag) – + + + – –
C-CBL(Q367P) (HA) – – + – + –
C-CBL(Y371S) (HA) – – – + – +
IP: Anti-EGFR
Blot: Anti-Ub
150 kDa
Anti-Flag
Anti-HA
b
CBL-B (Flag) – + + + – –
C-CBL(Q367P) (HA) – – + – + –
C-CBL(Y371S) (HA) – – – + – +
IP: Anti-EGFR
Blot: Anti-Ub
150 kDa
Anti-Flag
Anti-HA
c
Cell line V3MC BaF3/EPOR/JAK2 32D/FLT3
Cytokine SCF EPO FLT3LG
IP: Anti-c-KIT Anti-JAK2 Anti-FLT3
Mock
C-CBL(WT)
C-CBL(Q367P)
C-CBL(Y371S)
Blot: Anti-Ub
150 kDa
100 kDa
150 kDa
Blot:
Anti-c-KIT
Anti-JAK2
Anti-FLT3
d
Mock
C-CBL(WT)
C-CBL (Q367P)
C-CBL (Y371S)
0 10 60 0 10 60 0 10 60 0 10 60
150 kDa
Anti-p-EGFR
Anti-GAPDH
Mock
C-CBL(WT)
C-CBL (Q367P)
C-CBL (Y371S)
0 10 120 0 10 120 0 10 120 0 10 120
100 kDa
Anti-p-JAK2
Anti-α-tubulin

Figure 4

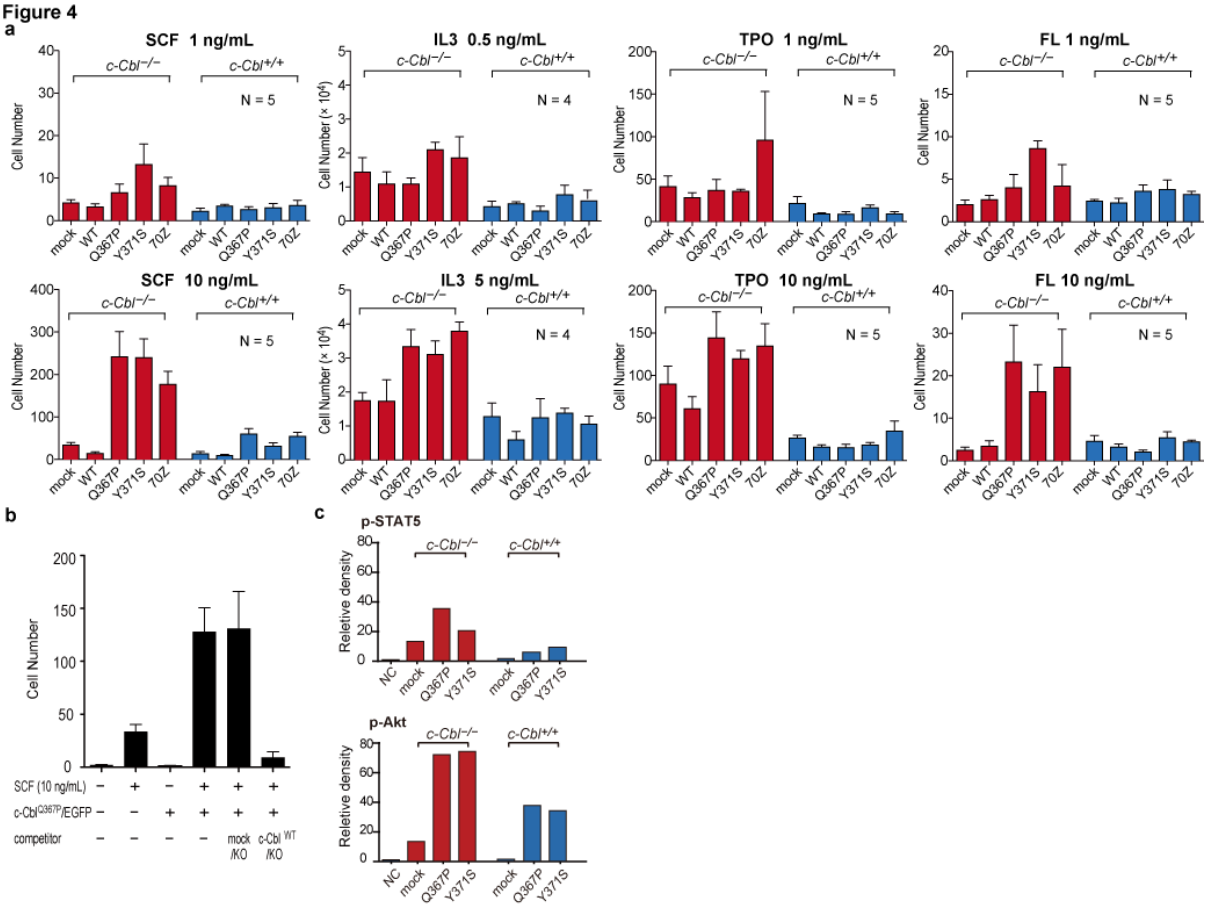

a
SCF 1 ng/mL
IL3 0.5 ng/mL
TPO 1 ng/mL
FL 1 ng/mL
SCF 10 ng/mL
IL3 5 ng/mL
TPO 10 ng/mL
FL 10 ng/mL
c-Cbl−/−
c-Cbl+/+
N = 5
N = 4
Cell Number
Cell Number (× 10⁴)
mock
WT
Q367P
Y371S
70Z
b
Cell Number
SCF (10 ng/mL)
c-Cbl Q367P/EGFP
competitor
mock /KO
c-Cbl WT /KO
c
p-STAT5
p-Akt
Relative density
NC
mock
Q367P
Y371S

**Figure Legends**

**Figure 1. Common UPD on 11q arm in MDS and *C-CBL* mutations.**

**a**, aUPD involving 11q arm was found in 17 cases with myeloid neoplasms, mostly CMML with normal karyotypes. Regions of CN gains, losses, and aUPD are depicted in different color as indicated. Histoloty of each cases are also shown by a color box. Asterisks denote *C-CBL*-mutated cases. **b**, CNAG output for MDS.U35. Both the total copy number (tCN) and allele-specific copy number (AsCN) plots show a focal copy number gain spanning a 1.4Mb segment within 3Mb of an 11q-aUPD region (upper left). Known genes and ESTs including *C-CBL* are presented below. A missense mutation (1250A > C, Y371S) was found in exon 8 of *C-CBL* in MDS.U35 (upper right). **c**, Comparison of amino acid sequences among human Cbl family proteins as well as their homologues in *C. elegance* (sli-1) and *drosophila* (D-Cbl). Numbering of amino acids is based on human C-CBL. Conserved amino acids are shown by shoadows. Positions of each mutated amino acid is indicated by asterisks, where heterozygous mutations are shown in red. **d**, Mutated amino acid positions in three-dimensional structure of a human C-CBL-UbcH7 complex. Mutated amino acids (red spheres) tend to be positioned within the Linker (yellow) and RING finger (green) domains, providing an interface for binding E2 conjugating enzymes, which may also affecting interaction with the TKB domain.

**Figure 2. Tumor suppressor functions of wild-type C-CBL**

**a**, Prolonged replating capacity of LSK cells transduced with mutant *C-CBL* (*C-CBL* $^{Q367P}$ and *C-CBL*$^{Y371S}$) compared with mock- or wild-type *C-CBL* transduced cells. Replating capacity in methylcellulose culture are shown as the mean number of colonies (+S.D.) per 1,000 replating cells at indicated times of replating. **b**, Increased spleen mass in *c-Cbl*$^{-/-}$ mice (red bar) compared to *c-Cbl*$^{+/+}$ mice (blue bar), shown as mean spleen weight +S.D. **c**, The mean numbers of total LSK (left panel) and CD34-negative LSK (right panel) cells (+S.D.) in bone marrow and/or spleen in *c-Cbl*$^{+/+}$ (blue bars) and

*c-Cbl*$^{-/-}$ mice (red bars). Bone marrow cells from bilateral tibias and femurs were counted for each mouse. **d**, Augmented colony forming potential of bone marrow cells from *c-Cbl*$^{-/-}$ mice (mean number of colonies per 5,000 bone marrow cells +S.D.). **e**, Kaplan-Meier survival curves of *c-Cbl*$^{+/+}$, *c-Cbl*$^{+/-}$, and *c-Cbl*$^{-/-}$ mice carrying a bcr/abl transgene, showing acceleration of blastic crisis in *c-Cbl*$^{+/-}$, and *c-Cbl*$^{-/-}$ mice. **f**, Wright–Giemsa staining (×600) of an enlarged lymph node in a *bcr/abl*$^{+}$ *c-Cbl*$^{-/-}$ mouse developing blastic crisis shows a massive infiltrate of immature leukemic blasts.

**Figure 3. Inhibitory actions of C-CBL mutants on wild-type C-CBL**

**a-b**, FLAG-tagged *C-CBL*$^{WT}$ or *CBL-B* were transfected into NIH3T3 cells stably transduced with hEGFR plus indicated HA-tagged-C-CBL mutants (C-CBL$^{Q367P}$ and C-CBL$^{Y371S}$). Anti-ubiquitin blots of immunoprecipitated hEGFR after EGF stimulation show the inhibitory action of the C-CBL mutants on the ubiquitinating activity of C-CBL (**a**) and CBL-B (**b**). Lower panels are anti-HA and anti-FLAG blots of total cell lysates. **c**, Effects of wild-type and mutant C-CBL on cytokine-induced ubiquitination of c-Kit, JAK2, and FLT3 in hematopoietic cells: V3MC, BaF3 cotransduced with human erythropoietin (hEpo) receptor and JAK2 (32D/EpoR/JAK2), and FLT3-transduced 32D (32D/FLT3), respectively. Each cell line was further transduced with indicated *C-CBL* mutants, and ubiquitination of immunoprecipitated kinases were detected by anti-ubiquitin blots at indicated time points after stimulation with SCF, erythropoietin, and FL. Anti-kinase blots of the precipitated kinases are shown below each panel. **d**, Phosphorylation of each kinase was also examined at indicated time points after the ligand stimulation using anti-phosphotyrosine (4G10) blots of the immunoprecipitated kinases. Precipitated kinases are shown below.

**Figure 4. Gain-of-functions of mutant C-CBL augmented by loss of wild-type C-CBL**

**a,** *c-Cbl*$^{+/+}$ and *c-Cbl*$^{-/-}$ LSK cells were transfected with various *C-CBL* IRES GFP

constructs, and 50 GFP-positive cells were sorted for *in vitro* serum-free culture containing different concentrations of SCF, IL3, TPO, and FL as indicated. The cell numbers were counted on day 5 and the mean cell numbers (+S.E.) are plotted. **b**, *c-Cbl*$^{-/-}$ LSK cells were cotransduced with *C-CBL*$^{Q367P}$-IRES-EGFP (*C-CBL*$^{Q367P}$/EGFP) and *mock*-IRES-Kusabira Orange (*mock*/KO) or *C-CBL*$^{WT}$-IRES-Kusabira Orange (*C-CBL*$^{WT}$/KO), and 50 GFP/KO double-positive cells were sorted into each well and subjected to cell proliferation assays in serum-free culture containing 10 ng/μL SCF. Mean cell numbers on day 5 (+S.E., N = 5) are plotted in the upper two panels. **c**, After 6 h of serum starvation, $1 \times 10^4$ of *c-Cbl*$^{+/+}$ and *c-Cbl*$^{-/-}$ LSK cells transduced with various C-CBL constructs were cultured in serum-free medium containing 10 ng/mL SCF and 10 ng/mL TPO for 15 min, and whole cell lysates were analyzed by immunoblotting using antibodies to STAT5, phosphorylated STAT5 (pSTAT5), Akt, and phosphorylated Akt (pAkt) (Supplementary Figure 14). The intensities of phosphorylated bands relative to total STAT5 (upper panel) and Akt (lower panel) are plotted.

## References


1 Knudson, A. G. Two genetic hits (more or less) to cancer. *Nature reviews. Cancer* **1**, 157-162, doi:10.1038/35101031 (2001).

2 James, C. *et al.* A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **434**, 1144-1148, doi:10.1038/nature03546 (2005).

3 Ryan, P. E., Davies, G. C., Nau, M. M. & Lipkowitz, S. Regulating the regulator: negative regulation of Cbl ubiquitin ligases. *Trends in biochemical sciences* **31**, 79-88, doi:10.1016/j.tibs.2005.12.004 (2006).

4 Schmidt, M. H. & Dikic, I. The Cbl interactome and its functions. *Nature reviews. Molecular cell biology* **6**, 907-918, doi:10.1038/nrm1762 (2005).

5 Thien, C. B. & Langdon, W. Y. Cbl: many adaptations to regulate protein tyrosine kinases. *Nature reviews. Molecular cell biology* **2**, 294-307, doi:10.1038/35067100 (2001).

6 Thien, C. B. & Langdon, W. Y. c-Cbl and Cbl-b ubiquitin ligases: substrate diversity and the negative regulation of signalling responses. *The Biochemical journal* **391**, 153-166, doi:10.1042/BJ20050892 (2005).

7 Corey, S. J. *et al.* Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nature reviews. Cancer* **7**, 118-129, doi:10.1038/nrc2047 (2007).

8 Jaffe, E., Harris, N., Stein, H., Vardiman, J. *World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues.*, (IARC press, 2002).

9 Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer research* **65**, 6071-6079, doi:10.1158/0008-5472.CAN-05-0465 (2005).

10 Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *American journal of human genetics* **81**, 114-126, doi:10.1086/518809 (2007).

11 Haase, D. Cytogenetic features in myelodysplastic syndromes. *Annals of hematology* **87**, 515-526, doi:10.1007/s00277-008-0483-y (2008).

12 Langdon, W. Y., Hartley, J. W., Klinken, S. P., Ruscetti, S. K. & Morse, H. C., 3rd. v-cbl, an oncogene from a dual-recombinant murine retrovirus that induces early B-lineage lymphomas. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 1168-1172

(1989).

13 Abbas, S., Rotmans, G., Lowenberg, B. & Valk, P. J. Exon 8 splice site mutations in the gene encoding the E3-ligase CBL are associated with core binding factor acute myeloid leukemias. *Haematologica* **93**, 1595-1597, doi:10.3324/haematol.13187 (2008).

14 Caligiuri, M. A. *et al.* Novel c-CBL and CBL-b ubiquitin ligase mutations in human acute myeloid leukemia. *Blood* **110**, 1022-1024, doi:10.1182/blood-2006-12-061176 (2007).

15 Sargin, B. *et al.* Flt3-dependent transformation by inactivating c-Cbl mutations in AML. *Blood* **110**, 1004-1012, doi:10.1182/blood-2007-01-066076 (2007).

16 Dunbar, A. J. *et al.* 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies. *Cancer research* **68**, 10349-10357, doi:10.1158/0008-5472.CAN-08-2754 (2008).

17 Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533-539 (2000).

18 Murphy, M. A. *et al.* Tissue hyperplasia and enhanced T-cell signalling via ZAP-70 in c-Cbl-deficient mice. *Molecular and cellular biology* **18**, 4872-4882 (1998).

19 Naramura, M., Kole, H. K., Hu, R. J. & Gu, H. Altered thymic positive selection and intracellular signals in Cbl-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 15547-15552 (1998).

20 Rathinam, C., Thien, C. B., Langdon, W. Y., Gu, H. & Flavell, R. A. The E3 ubiquitin ligase c-Cbl restricts development and functions of hematopoietic stem cells. *Genes & development* **22**, 992-997, doi:10.1101/gad.1651408 (2008).

21 Honda, H. *et al.* Acquired loss of p53 induces blastic transformation in p210(bcr/abl)-expressing hematopoietic cells: a transgenic study for blast crisis of human CML. *Blood* **95**, 1144-1150 (2000).

22 Zeng, S., Xu, Z., Lipkowitz, S. & Longley, J. B. Regulation of stem cell factor receptor signaling by Cbl family proteins (Cbl-b/c-Cbl). *Blood* **105**, 226-232, doi:10.1182/blood-2004-05-1768 (2005).

23 Kaushansky, K. Hematopoietic growth factors, signaling and the chronic myeloproliferative disorders. *Cytokine & growth factor reviews* **17**, 423-430, doi:10.1016/j.cytogfr.2006.09.005 (2006).

24 Naramura, M. *et al.* c-Cbl and Cbl-b regulate T cell responsiveness by promoting ligand-induced TCR down-modulation. *Nature immunology* **3**, 1192-1199, doi:10.1038/ni855 (2002).

25 Dittmer, D. *et al.* Gain of function mutations in p53. *Nature genetics* **4**, 42-46, doi:10.1038/ng0593-42 (1993).

26 Lang, G. A. *et al.* Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* **119**, 861-872, doi:10.1016/j.cell.2004.11.006 (2004).

27 Finlay, C. A., Hinds, P. W. & Levine, A. J. The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57**, 1083-1093 (1989).

28 Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971-974, doi:10.1038/nature07399 (2008).

29 Blake, T. J., Shapiro, M., Morse, H. C., 3rd & Langdon, W. Y. The sequences of the human and mouse c-cbl proto-oncogenes show v-cbl was generated by a large truncation encompassing a proline-rich domain and a leucine zipper-like motif. *Oncogene* **6**, 653-657 (1991).

30 Hamanaka, S. *et al.* Stable transgene expression in mice generated from retrovirally transduced embryonic stem cells. *Molecular therapy : the journal of the American Society of Gene Therapy* **15**, 560-565, doi:10.1038/sj.mt.6300063 (2007).

31 Sanuki, S. *et al.* A new red fluorescent protein that allows efficient marking of murine hematopoietic stem cells. *The journal of gene medicine* **10**, 965-971, doi:10.1002/jgm.1232 (2008).

32 Ory, D. S., Neugeboren, B. A. & Mulligan, R. C. A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 11400-11406 (1996).

33 Suzuki, A. *et al.* Feasibility of ex vivo gene therapy for neurological disorders using the new retroviral vector GCDNsap packaged in the vesicular stomatitis virus G protein. *Journal of neurochemistry* **82**, 953-960 (2002).

34 Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242-245 (1996).

**Supplementary Information**

# Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms

Masashi Sanada, Takahiro Suzuki, Lee-Yung Shih, Makoto Otsu, Satoshi Yamazaki, Motohiro Kato, Azusa Tamura, Hiroaki Honda, Mamiko Sakata-Yanagimoto, Keiki Kumano, Hideaki Oda, Tetsuya Yamagata, Junko Takita, Noriko Gotoh, Kumi Nakazaki, Norihiko Kawamata, Masafumi Onodera, Masaharu Nobuyoshi, Yasuhide Hayashi, Hiroshi Harada, Mineo Kurokawa, Shigeru Chiba, Hiraku Mori, Keiya Ozawa, Mitsuhiro Omine, Hisamaru Hirai, Hiromitsu Nakauchi, H. Phillip Koeffler, Seishi Ogawa

**Contents**

**Supplementary Tables**
Table 1. Demographic features of 222 cases with myeloid neoplasms
Table 2. Clinical information of individual samples with myeloid neoplasms subjected to SNP arrays/mutation studies
Table 3. Comparison of sensitivity of detecting genetic lesions between conventional cytogenetics and SNP array analysis using CNAG/AsCNAR in 222 patients with myeloid neoplasms
Table 4. Sensitivity of detecting aUPD in different algorithms
Table 5. Positions of aUPD found in 222 cases with myeloid neoplasms
Table 6. Summary of 11q-aUPD and *c–Cbl* mutations
Table 7. Characteristics of myeloid neoplasms either with or without *c–Cbl* mutations
Table 8. Primer sets and their PCR conditions used in this study
Table 9. Antibodies used in this study

**Supplementary Figures**
Figure 1. Integrated pictures of copy number alterations and LOH on ideograms in myeloid neoplasms
Figure 2. Sensitive detection of copy number neutral LOH (or aUPD) by AsCNAR
Figure 3. Shared identical-by-descent (IBD) alleles by parents versus aUPD
Figure 4. Chromosomal distribution of aUPD in cases with or without copy number abnormalities
Figure 5. Distribution of genetic lesions in 222 samples of myeloid neoplasms
Figure 6. aUPD in 11q arm and/or *c–Cbl* mutations in myeloid neoplasms
Figure 7. Effect of *c-Cbl* mutation status on overall survival in CMML cases
Figure 8. Oncogenic potentials of mutant c-Cbl
Figure 9. Activation of Akt in NIH3T3 transduced with c-Cbl mutants
Figure 10. Generation of *c–Cbl* null mice by gene targeting
Figure 11. Flow cytometric analysis of bone marrow cells and spleen cells from $c–Cbl^{+/+}$ and $c–Cbl^{-/-}$ mice
Figure 12. Immunoblot analysis of expression of c-Cbl and Cbl-b in LSK HSPCs
Figure 13. Assay for sensitivity to SCF in LSK HSPCs transduced with c-Cbl mutants
Figure 14. Immunoblot analysis of activation of STAT5 and Akt in LSK HSPCs stimulated with SCF + TPO

**References**

**Supplementary Table 1. Demographic features of 222 cases with myeloid neoplasms**

| | | | |
|---|---|---|---|
| **Age** | 22–91 (median 67) | | |
| **Sex** | Female | 71 | 32% |
| | Male | 151 | 68% |
| **WHO Classification** | | | |
| **MDS** | | | |
| | RCUD/RARS | 4 | 2% |
| | RCMD | 25 | 11% |
| | MDS associated with isolated del(5q) | 6 | 3% |
| | RAEB–1 | 27 | 12% |
| | RAEB–2 | 42 | 19% |
| | MDSu | 4 | 2% |
| **MDS/MPN** | | | |
| | CMML–1 | 54 | 24% |
| | CMML–2 | 31 | 14% |
| **AML** | | | |
| | AML with myelodysplasia-related changes¶ | 24 | 11% |
| | Therapy-related myeloid neoplasms | 4 | 2% |
| **IPSS†** | | | |
| | Low | 12 | 7% |
| | Int–1 | 54 | 32% |
| | Int–2 | 41 | 24% |
| | High | 35 | 20% |
| **Cytogenetics** | | | |
| | normal karyotypes | 75 | 43% |
| | –5/5q– | 20 | 12% |
| | –7/7q– | 22 | 13% |
| | +8 | 15 | 9% |
| | del(20q) | 10 | 6% |
| | complex(≥5) | 9 | 5% |
| | unknown‡ | 29 | 18% |

† Classification in the International Prognostic Scoring System
‡ Including cases, in which no metaphases were obtained
¶ Including 18 cases transformed from MDS and 6 cases with RAEB-t in FAB classification

**Supplementary Table 2. Clinical information of individual samples with myeloid neoplasms subjected to SNP arrays/mutation studies**

| Sample ID | Array type | Sex | Age | Ethnicity | FAB† type | WHO classification¶ | IPSS‡ |
|---|---|---|---|---|---|---|---|
| MDS.U1 | 50K *Xba*I | F | 62 | Chinese | RA | RCMD | LOW |
| MDS.U2 | 50K *Xba*I | M | 68 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U3 | 50K *Xba*I | F | 55 | Chinese | CMML | CMML2 | HIGH |
| MDS.U5 | 50K *Xba*I | F | 71 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U6 | 50K *Xba*I | M | 42 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U7 | 50K *Xba*I | M | 71 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U8 | 50K *Xba*I | F | 65 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U9 | 50K *Xba*I | F | 48 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U10 | 50K *Xba*I | M | 50 | Chinese | RA | RCMD | INT–1 |
| MDS.U11 | 50K *Xba*I | M | 80 | Chinese | RARS | RCMD | INT–2 |
| MDS.U12 | 50K *Xba*I | M | 45 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U14 | 50K *Xba*I | M | 42 | Chinese | RA | RCMD | INT–1 |
| MDS.U15 | 50K *Xba*I | M | 76 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U16 | 50K *Xba*I | M | 68 | Chinese | RAEB | RAEB1 | INT–2 |
| MDS.U17 | 50K *Xba*I | F | 80 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U18 | 50K *Xba*I | F | 67 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U19 | 50K *Xba*I | M | 68 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U20 | 50K *Xba*I | F | 45 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U22 | 50K *Xba*I | F | 22 | Chinese | RA | RCMD | INT–1 |
| MDS.U23 | 50K *Xba*I | M | 67 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U24 | 250K *Nsp*I | F | 65 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U25 | 50K *Xba*I | M | 41 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U26 | 50K *Xba*I | M | 56 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U27 | 50K *Xba*I | F | 86 | Chinese | CMML | CMML1 | LOW |
| MDS.U28 | 50K *Xba*I | M | 64 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U29 | 50K *Xba*I | M | 62 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U31 | 50K *Xba*I | F | 64 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U32 | 50K *Xba*I | F | 68 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U33 | 50K *Xba*I | F | 67 | Chinese | CMML | CMML2 | HIGH |
| MDS.U34 | 50K *Xba*I | M | 64 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U35 | 50K *Xba*I | F | 70 | Chinese | RAEB | RAEB1 | INT–2 |
| MDS.U36 | 50K *Xba*I | M | 65 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U37 | 50K *Xba*I | M | 67 | Chinese | RA | RCMD | LOW |
| MDS.U38 | 50K *Xba*I | M | 45 | Chinese | RA | RCMD | INT–1 |
| MDS.U40 | 50K *Xba*I | M | 78 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U41 | 50K *Xba*I | F | 73 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U42 | 50K *Xba*I | M | 75 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U43 | 50K *Xba*I | F | 38 | Chinese | RA | RCMD | INT–1 |
| MDS.U44 | 50K *Xba*I | M | 57 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U45 | 50K *Xba*I | M | 34 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U46 | 50K *Xba*I | M | 78 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U47 | 50K *Xba*I | M | 77 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U48 | 50K *Xba*I | M | 66 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U49 | 50K *Xba*I | M | 58 | Chinese | CMML | CMML1 | INT–2 |
| MDS.U50 | 50K *Xba*I | M | 74 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U51 | 50K *Xba*I | F | 73 | Chinese | MDS | MDS–u | LOW |
| MDS.U54 | 50K *Xba*I | F | 77 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U56 | 50K *Xba*I | M | 88 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U57 | 50K *Xba*I | M | 70 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U58 | 50K *Xba*I | M | 73 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U59 | 50K *Xba*I | M | 57 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U60 | 50K *Xba*I | M | 84 | Chinese | RAEB | RAEB1 | INT–2 |
| MDS.U61 | 50K *Xba*I | M | 66 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U62 | 50K *Xba*I | M | 65 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U63 | 50K *Xba*I | F | 59 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U64 | 50K *Xba*I | M | 73 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U65 | 50K *Xba*I | F | 72 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U66 | 50K *Xba*I | M | 78 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U67 | 50K *Xba*I | M | 51 | Chinese | RAEB | RAEB2 | INT–2 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MDS.U68 | 50K *Xba*I | M | 82 | Chinese | CMML | CMML2 | HIGH |
| MDS.U69 | 50K *Xba*I | M | 70 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U70 | 50K *Xba*I | M | 80 | Chinese | CMML | CMML2 | HIGH |
| MDS.U71 | 50K *Xba*I | F | 62 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U72 | 50K *Xba*I | F | 57 | Chinese | RAEB | RAEB2 | INT–2 |
| MDS.U73 | 50K *Xba*I | M | 76 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U74 | 50K *Xba*I | M | 72 | Chinese | RAEB | RAEB2 | HIGH |
| MDS.U75 | 50K *Xba*I | M | 70 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U76 | 50K *Xba*I | M | 74 | Chinese | RAEB | RAEB1 | INT–1 |
| MDS.U77 | 250K *Nsp*I | M | 76 | Chinese | CMML | CMML1 | LOW |
| MDS.U101 | 50K *Hin*dIII | M | ND | Japanese | MDS | MDS–u | ND |
| MDS.U102 | 50K *Hin*dIII | M | 62 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U103 | 50K *Hin*dIII | M | 67 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U104 | 50K *Hin*dIII | M | 57 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U105 | 50K *Hin*dIII | M | 56 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U106 | 50K *Hin*dIII | M | 72 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U107 | 50K *Xba*I | M | 50 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U108 | 50K *Hin*dIII | F | ND | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U109 | 50K *Hin*dIII | M | ND | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U110 | 50K *Hin*dIII | M | 81 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U111 | 50K *Hin*dIII | F | ND | Japanese | RA | 5q– syndrome | ND |
| MDS.U112 | 50K *Hin*dIII | F | 58 | Japanese | RAEBt | RAEB2 | ND |
| MDS.U113 | 50K *Hin*dIII | F | 55 | Japanese | RAEBt | MDS-related AML | HIGH |
| MDS.U114 | 50K *Hin*dIII | M | 72 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U115 | 50K *Hin*dIII | M | ND | Japanese | CMML | CMML1 | ND |
| MDS.U116 | 50K *Hin*dIII | M | 68 | Japanese | RAEBt | RAEB2 | INT–2 |
| MDS.U117 | 50K *Hin*dIII | M | ND | Japanese | MDS | MDS–u | ND |
| MDS.U118 | 50K *Hin*dIII | F | 89 | Japanese | CMML | CMML1 | LOW |
| MDS.U119 | 50K *Hin*dIII | F | 68 | Japanese | RAEB | t-MN(RAEB2) | HIGH |
| MDS.U120 | 50K *Hin*dIII | F | 71 | Japanese | RA | RCMD | LOW |
| MDS.U121 | 50K *Hin*dIII | F | 32 | Japanese | RA | RCUD | INT–1 |
| MDS.U122 | 50K *Hin*dIII | M | 74 | Japanese | RA | RCMD | INT–1 |
| MDS.U123 | 50K *Hin*dIII | M | 78 | Japanese | RAEBt | RAEB2 | HIGH |
| MDS.U124 | 50K *Xba*I | M | ND | Caucasian | RA | 5q– syndrome | ND |
| MDS.U125 | 50K *Xba*I | M | ND | Caucasian | MDS | MDS–u | ND |
| MDS.U126 | 50K *Hin*dIII | F | 79 | Japanese | RA | RCMD | LOW |
| MDS.U127 | 50K *Hin*dIII | F | 72 | Japanese | RA | RCMD | LOW |
| MDS.U128 | 50K *Hin*dIII | F | 91 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U129 | 50K *Hin*dIII | M | 42 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U130 | 50K *Hin*dIII | M | 47 | Japanese | CMML | CMML2 | HIGH |
| MDS.U131 | 50K *Hin*dIII | M | 69 | Japanese | RAEB | RAEB2 | HIGH |
| MDS.U132 | 50K *Hin*dIII | M | 36 | Japanese | RAEBt | MDS-related AML | HIGH |
| MDS.U133 | 250K *Nsp*I | M | 43 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U134 | 250K *Nsp*I | M | 34 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U135 | 250K *Nsp*I | M | 73 | Japanese | RA | RCMD | INT–1 |
| MDS.U136 | 250K *Nsp*I | M | 67 | Japanese | RA | RCMD | INT–1 |
| MDS.U137 | 250K *Nsp*I | M | 71 | Japanese | RA | RCMD | INT–1 |
| MDS.U138 | 250K *Nsp*I | M | 64 | Japanese | RAEB | RAEB2 | INT–1 |
| MDS.U139 | 250K *Nsp*I | M | 69 | Japanese | CMML | CMML1 | INT–1 |
| MDS.U140 | 50K *Hin*dIII | M | 62 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U141 | 250K *Nsp*I | F | 68 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U142 | 250K *Nsp*I | M | 74 | Japanese | CMML | CMML–2 | INT–2 |
| MDS.U143 | 250K *Nsp*I | M | 72 | Japanese | RAEB | RAEB1 | INT–1 |
| MDS.U144 | 250K *Nsp*I | M | 66 | Japanese | RA | RCMD | INT–1 |
| MDS.U145 | 250K *Nsp*I | F | 55 | Japanese | RARS | RARS | LOW |
| MDS.U146 | 250K *Nsp*I | M | 24 | Japanese | RAEB | RAEB2 | INT–2 |
| MDS.U147 | 250K *Nsp*I | F | 35 | Japanese | RAEB | RAEB2 | HIGH |
| MDS.U148 | 250K *Nsp*I | M | 39 | Japanese | RAEB | RAEB1 | INT–2 |
| MDS.U149 | 250K *Nsp*I | M | 63 | Japanese | RAEB | RAEB1 | INT–2 |
| MDS.U150 | 250K *Nsp*I | M | 76 | Japanese | RAEBt | RAEB2 | HIGH |
| MDS.U151 | 50K *Hin*dIII | M | 58 | Japanese | MDS–derived AML | MDS-related AML | NA |
| MDS.U152 | 50K *Hin*dIII | F | ND | Caucasian | RA | 5q– syndrome | ND |
| MDS.U153 | 50K *Hin*dIII | F | ND | Caucasian | RA | 5q– syndrome | ND |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MDS.U154 | 50K *Hin*dIII | F | ND | Caucasian | RA | 5q– syndrome | ND |
| MDS.U155 | 50K *Hin*dIII | F | ND | Caucasian | RA | 5q– syndrome | ND |
| MDS.U156 | 50K *Hin*dIII | M | 57 | Japanese | RAEB | RAEB1 | INT–1 |
| MDS.U157 | 50K *Hin*dIII | F | 78 | Japanese | RAEBt | RAEB2 | INT–2 |
| MDS.U158 | 50K *Hin*dIII | M | 75 | Japanese | RA | RCMD | INT–1 |
| MDS.U159 | 50K *Hin*dIII | M | 67 | Japanese | RAEB | RAEB2 | INT–2 |
| MDS.U160 | 50K *Hin*dIII | M | 69 | Japanese | RAEB | RAEB1 | INT–2 |
| MDS.U161 | 50K *Hin*dIII | M | 35 | Japanese | CMML | CMML1 | INT–1 |
| MDS.U162 | 50K *Hin*dIII | F | 73 | Japanese | RAEBt | RAEB2 | INT–1 |
| MDS.U163 | 50K *Hin*dIII | M | 52 | Japanese | RAEBt | RAEB2 | INT–2 |
| MDS.U164 | 50K *Hin*dIII | F | 64 | Japanese | CMML | CMML2 | HIGH |
| MDS.U165 | 50K *Hin*dIII | M | 29 | Japanese | RA | RCMD | INT–1 |
| MDS.U166 | 50K *Hin*dIII | M | ND | Japanese | RAEBt | MDS-related AML | HIGH |
| MDS.U167 | 50K *Hin*dIII | F | 78 | Japanese | RAEBt | MDS-related AML | INT–2 |
| MDS.U168 | 50K *Hin*dIII | M | 53 | Japanese | RA | RCMD | INT–2 |
| MDS.U169 | 50K *Xba*I | M | 78 | Japanese | RAEB | t-MN(RAEB1) | INT–2 |
| MDS.U170 | 50K *Hin*dIII | F | 68 | Japanese | CMML | CMML2 | INT–2 |
| MDS.U171 | 50K *Hin*dIII | M | 61 | Japanese | RA | RCMD | INT–1 |
| MDS.U172 | 50K *Hin*dIII | M | 79 | Japanese | RA | RCMD | LOW |
| MDS.U173 | 50K *Hin*dIII | M | 65 | Japanese | CMML | CMML1 | INT–2 |
| MDS.U174 | 50K *Hin*dIII | F | 55 | Japanese | RAEBt | RAEB2 | INT–2 |
| MDS.U175 | 50K *Hin*dIII | M | 58 | Japanese | RAEB | RAEB1 | INT–1 |
| MDS.U176 | 50K *Hin*dIII | M | 63 | Japanese | RAEBt | RAEB2 | INT–2 |
| MDS.U177 | 50K *Hin*dIII | F | 75 | Japanese | RAEBt | RAEB1 | INT–1 |
| MDS.U178 | 50K *Hin*dIII | F | 74 | Japanese | RA | RCMD | INT–1 |
| MDS.U179 | 50K *Hin*dIII | M | 70 | Japanese | RAEB | RAEB1 | INT–1 |
| MDS.U180 | 50K *Hin*dIII | M | 87 | Japanese | RAEBt | RAEB2 | INT–2 |
| MDS.U181 | 50K *Xba*I | M | 82 | Japanese | RA | RCMD | INT–1 |
| MDS.U182 | 50K *Hin*dIII | F | 47 | Japanese | RAEB | RAEB1 | INT–2 |
| MDS.U183 | 50K *Hin*dIII | F | 74 | Japanese | RAEB | RAEB1 | HIGH |
| MDS.U184 | 50K *Xba*I | M | 78 | Japanese | RAEBt | MDS-related AML | HIGH |
| MDS.U185 | 50K *Hin*dIII | M | 76 | Japanese | CMML | CMML2 | HIGH |
| MDS.U186 | 50K *Hin*dIII | M | 53 | Japanese | RAEB | RAEB2 | HIGH |
| MDS.U187 | 50K *Hin*dIII | M | 59 | Japanese | RAEB | RAEB2 | HIGH |
| MDS.U188 | 50K *Hin*dIII | M | 54 | Japanese | RAEB | t-MN(RAEB2) | HIGH |
| MDS.U189 | 50K *Hin*dIII | M | 64 | Japanese | RAEBt | MDS-related AML | HIGH |
| MDS.U190 | 50K *Hin*dIII | M | 67 | Japanese | RA | RCMD | INT–1 |
| MDS.U191 | 50K *Hin*dIII | M | 72 | Japanese | CMML | CMML2 | INT–1 |
| MDS.U192 | 50K *Hin*dIII | M | 70 | Japanese | RAEBt | RAEB2 | HIGH |
| MDS.U193 | 50K *Xba*I | M | 71 | Japanese | MDS-derived AML | MDS-related AML | NA |
| MDS.U194 | 50K *Xba*I | M | 77 | Japanese | RAEB | RAEB2 | INT–2 |
| MDS.U195 | 50K *Hin*dIII | F | 74 | Japanese | RAEB | RAEB2 | INT–2 |
| MDS.U196 | 50K *Hin*dIII | M | 75 | Japanese | RAEB | t-MN(RAEB2) | HIGH |
| MDS.U197 | 50K *Hin*dIII | M | 70 | Japanese | RA | RCUD | INT–1 |
| MDS.U198 | 50K *Hin*dIII | F | 58 | Japanese | RA | RCUD | INT–1 |
| MDS.U199 | 50K *Hin*dIII | M | 76 | Japanese | CMML | CMML1 | LOW |
| MDS.U200 | 50K *Hin*dIII | M | 44 | Japanese | RA | RCMD | INT–1 |
| MDS.U201 | 50K *Hin*dIII | M | 63 | Japanese | RAEBt | RAEB2 | INT–2 |
| MDS.U202 | 50K *Hin*dIII | M | 73 | Japanese | RAEBt | RAEB2 | HIGH |
| MDS.U302 | 250K *Nsp*I | M | 63 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U303 | 250K *Nsp*I | F | 59 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U304 | 250K *Nsp*I | F | 61 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U305 | 250K *Nsp*I | F | 83 | Chinese | CMML | CMML2 | HIGH |
| MDS.U306 | 250K *Nsp*I | M | 30 | Chinese | RAEB-T | CMML2 | HIGH |
| MDS.U307 | 250K *Nsp*I | M | 59 | Chinese | CMML | CMML1 | LOW/INT-1 |
| MDS.U308 | 250K *Nsp*I | F | 61 | Chinese | RAEB-T | CMML2 | INT–1 |
| MDS.U309 | 250K *Nsp*I | F | 74 | Chinese | RAEB-T | CMML2 | INT-2/HIGH |
| MDS.U310 | 250K *Nsp*I | M | 67 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U311 | 250K *Nsp*I | M | 90 | Chinese | RAEB-T | CMML2 | INT-1/INT-2 |
| MDS.U312 | 250K *Nsp*I | M | 71 | Chinese | RAEB-T | CMML2 | INT-1/INT-2 |
| MDS.U313 | 250K *Nsp*I | M | 78 | Chinese | RAEB-T | CMML2 | INT-2/HIGH |
| MDS.U314 | 250K *Nsp*I | F | 82 | Chinese | CMML | CMML1 | LOW/INT-1 |
| MDS.U315 | 250K *Nsp*I | F | 32 | Chinese | CMML | CMML2 | INT–2 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| MDS.U316 | 250K *Nsp*I | F | 40 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U317 | 250K *Nsp*I | M | 81 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U318 | 250K *Nsp*I | M | 77 | Chinese | CMML | CMML1 | INT–2 |
| MDS.U319 | 250K *Nsp*I | F | 77 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U320 | 250K *Nsp*I | F | 73 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U321 | 250K *Nsp*I | F | 67 | Chinese | CMML | CMML2 | INT-2/HIGH |
| MDS.U322 | 250K *Nsp*I | M | 79 | Chinese | CMML | CMML1 | LOW |
| MDS.U323 | 250K *Nsp*I | F | 80 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U324 | 250K *Nsp*I | M | 77 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U325 | 250K *Nsp*I | F | 67 | Chinese | CMML | CMML1 | LOW |
| MDS.U326 | 250K *Nsp*I | M | 80 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U327 | 250K *Nsp*I | M | 41 | Chinese | CMML | CMML2 | INT-1/INT-2 |
| MDS.U328 | 250K *Nsp*I | F | 80 | Chinese | CMML | CMML1 | LOW |
| MDS.U329 | 250K *Nsp*I | M | 64 | Chinese | CMML | CMML1 | INT–2 |
| MDS.U330 | 250K *Nsp*I | M | 79 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U331 | 250K *Nsp*I | M | 72 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U332 | 250K *Nsp*I | M | 74 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U333 | 250K *Nsp*I | M | 76 | Chinese | CMML | CMML2 | HIGH |
| MDS.U334 | 250K *Nsp*I | F | 70 | Chinese | CMML | CMML1 | LOW/INT-1 |
| MDS.U335 | 250K *Nsp*I | M | 71 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U336 | 250K *Nsp*I | M | 86 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U337 | 250K *Nsp*I | M | 60 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U338 | 250K *Nsp*I | M | 79 | Chinese | CMML | CMML1 | ND |
| MDS.U339 | 250K *Nsp*I | M | 77 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U340 | 250K *Nsp*I | F | 84 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U341 | 250K *Nsp*I | M | 56 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U342 | 250K *Nsp*I | F | 77 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U343 | 250K *Nsp*I | M | 39 | Chinese | CMML | CMML1 | LOW/INT-1 |
| MDS.U344 | 250K *Nsp*I | M | 82 | Chinese | CMML | CMML1 | INT-1/INT-2 |
| MDS.U345 | 250K *Nsp*I | M | 54 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U346 | 250K *Nsp*I | M | 75 | Chinese | CMML | CMML | LOW |
| MDS.U347 | 250K *Nsp*I | M | 79 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U348 | 250K *Nsp*I | M | 71 | Chinese | CMML | CMML1 | INT–1 |
| MDS.U349 | 250K *Nsp*I | M | 83 | Chinese | CMML | CMML1 | LOW |
| MDS.U350 | 250K *Nsp*I | M | 88 | Chinese | CMML | CMML2 | INT–2 |
| MDS.U351 | 250K *Nsp*I | M | 69 | Chinese | CMML | CMML1 | INT–1 |

† The French–American–British Classification

‡ The International Prognostic Scoring System; NA: not applicable, ND: not determined

¶ t-MN:Therapy-related myeloid neoplasms; MDS-related AML:AML with MDS-related changes; 5q-syndrome:MDS with isolated del(5q); MDSu:MDS–unclassified

**Supplementary Table 3. Comparison of sensitivity of detecting genetic lesions between conventional cytogenetics and SNP array analysis using CNAG/AsCNAR in 222 patients with myeloid neoplasms**

| Location of lesion | | Cytogenetics (N=171) | | SNP array using CNAG/AsCNAR | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | All (N=171) | | Normal cytogenetics (N=74) | | Unknown cytogenetics (N=29)† | |
| | | No.(+)‡ | % | No.(+) | % | No.(+) | % | No.(+) | % |
| Chromosome | 1 | 9 | 5.3 | 21 | 12.3 | 5 | 6.7 | 1 | 1.3 |
| | 2 | 2 | 1.2 | 8 | 4.7 | 0 | 0.0 | 0 | 0.0 |
| | 3 | 4 | 2.3 | 14 | 8.2 | 2 | 2.7 | 3 | 4.0 |
| | 4 | 3 | 1.8 | 10 | 5.8 | 1 | 1.3 | 2 | 2.7 |
| | 5 | 20 | 11.7 | 41 | 24.0 | 3 | 4.0 | 8 | 10.7 |
| | 6 | 6 | 3.5 | 17 | 9.9 | 3 | 4.0 | 3 | 4.0 |
| | 7 | 23 | 13.5 | 41 | 24.0 | 3 | 4.0 | 9 | 12.0 |
| | 8 | 18 | 10.5 | 27 | 15.8 | 3 | 4.0 | 4 | 5.3 |
| | 9 | 8 | 4.7 | 20 | 11.7 | 0 | 0.0 | 5 | 6.7 |
| | 10 | 3 | 1.8 | 6 | 3.5 | 2 | 2.7 | 2 | 2.7 |
| | 11 | 3 | 1.8 | 16 | 9.4 | 1 | 1.3 | 4 | 5.3 |
| | 12 | 5 | 2.9 | 20 | 11.7 | 2 | 2.7 | 3 | 4.0 |
| | 13 | 9 | 5.3 | 15 | 8.8 | 2 | 2.7 | 3 | 4.0 |
| | 14 | 2 | 1.2 | 3 | 1.8 | 1 | 1.3 | 1 | 1.3 |
| | 15 | 2 | 1.2 | 6 | 3.5 | 1 | 1.3 | 0 | 0.0 |
| | 16 | 4 | 2.3 | 10 | 5.8 | 0 | 0.0 | 1 | 1.3 |
| | 17 | 5 | 2.9 | 21 | 12.3 | 3 | 4.0 | 4 | 5.3 |
| | 18 | 6 | 3.5 | 13 | 7.6 | 0 | 0.0 | 2 | 2.7 |
| | 19 | 5 | 2.9 | 12 | 7.0 | 0 | 0.0 | 3 | 4.0 |
| | 20 | 10 | 5.8 | 16 | 9.4 | 0 | 0.0 | 2 | 2.7 |
| | 21 | 6 | 3.5 | 11 | 6.4 | 2 | 2.7 | 3 | 4.0 |
| | 22 | 3 | 1.8 | 3 | 1.8 | 0 | 0.0 | 0 | 0.0 |
| Balanced translocation¶ | | 3 | 1.8 | NA | NA | NA | NA | NA | NA |
| Unknown§ | | 29 | 16.9 | NA | NA | NA | NA | NA | NA |
| aUPD¶ | | NA | NA | 51 | 29.8 | 24 | 32.0 | 6 | 20.7 |
| no abnormality§ | | 74 | 43.3 | 48 | 28.1 | 31 | 41.3 | 14 | 18.7 |

NA: not applicable

† including cases in which no metaphases were obtained

‡ indicating presence of abnormality

¶ number of cases having balanced translocations and aUPDs in any chromosomes

¶ number of cases having no abnormalities or cases whose cytogenetic profile was unknown

**Supplementary Table 4. Sensitivity of detecting aUPD in different algorithms**

| Chr | SNP call–based algorithm | AsCNAR | SNP call–based algorithm only† | AsCNAR–only‡ |
|---|---|---|---|---|
| 1 | 6 | 8 | 0 | 2 |
| 2 | 0 | 1 | 0 | 1 |
| 3 | 0 | 1 | 0 | 1 |
| 4 | 7 | 9 | 0 | 2 |
| 5 | 0 | 0 | 0 | 0 |
| 6 | 0 | 1 | 0 | 1 |
| 7 | 5 | 9 | 0 | 4 |
| 8 | 2 | 2 | 0 | 0 |
| 9 | 2 | 2 | 0 | 0 |
| 10 | 1 | 1 | 0 | 0 |
| 11 | 9 | 21 | 0 | 12 |
| 12 | 1 | 2 | 0 | 1 |
| 13 | 1 | 2 | 0 | 1 |
| 14 | 2 | 6 | 0 | 4 |
| 15 | 0 | 0 | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 |
| 17 | 8 | 12 | 0 | 4 |
| 18 | 2 | 2 | 0 | 0 |
| 19 | 2 | 3 | 0 | 1 |
| 20 | 0 | 0 | 0 | 0 |
| 21 | 4 | 4 | 0 | 0 |
| 22 | 1 | 1 | 0 | 0 |
| Total | 53 | 86 | 0 | 33 |

† detected only by SNP call–based algorithm
‡ detected only by AsCNAR

**Supplementary Table 5. Positions of aUPD found in 222 cases with myeloid neoplasms**

| Sample ID | Array | Chromosome | StartPos | EndPos | Length(Mb) |
|---|---|---|---|---|---|
| MDS.U63 | 50K_Xba | 1 | pter | 52,530,209 | 52 |
| MDS.U170 | 50K_Hind | 1 | pter | 53,910,583 | 52 |
| MDS.U153 | 50K_Hind | 1 | pter | 86,605,785 | 84 |
| MDS.U34 | 50K_Xba | 1 | pter | 143,896,763 | 143 |
| MDS.U113 | 50K_Hind | 1 | 142,397,633 | qter | 102 |
| MDS.U155 | 50K_Hind | 1 | 144,207,899 | qter | 101 |
| MDS.U193 | 50K_Xba | 1 | 149,130,054 | qter | 96 |
| MDS.U65 | 50K_Xba | 1 | 184,479,444 | qter | 61 |
| MDS.U331 | 250K_Nsp | 2 | pter | 88,578,906 | 89 |
| MDS.U328 | 250K_Nsp | 3 | 124,210,237 | qter | 75 |
| MDS.U335 | 250K_Nsp | 4 | 48,221,993 | qter | 143 |
| MDS.U339 | 250K_Nsp | 4 | 48,324,858 | qter | 143 |
| MDS.U349 | 250K_Nsp | 4 | 75,102,506 | qter | 116 |
| MDS.U56 | 50K_Xba | 4 | 53,339,597 | qter | 138 |
| MDS.U75 | 50K_Xba | 4 | 54,076,291 | qter | 137 |
| MDS.U160 | 50K_Hind | 4 | 77,610,642 | qter | 113 |
| MDS.U46 | 50K_Xba | 4 | 82,667,092 | qter | 108 |
| MDS.U74 | 50K_Xba | 4 | 89,193,129 | qter | 102 |
| MDS.U32 | 50K_Xba | 4 | 92,097,728 | qter | 99 |
| MDS.U27 | 50K_Xba | 6 | pter | 26,138,471 | 26 |
| MDS.U34 | 50K_Xba | 7 | 57,653,125 | qter | 101 |
| MDS.U32 | 50K_Xba | 7 | 63,770,059 | qter | 93 |
| MDS.U150 | 250K_Nsp | 7 | 67,300,475 | qter | 91 |
| MDS.U1 | 50K_Xba | 7 | 77,561,689 | qter | 81 |
| MDS.U351 | 250K_Nsp | 7 | 91,790,601 | qter | 67 |
| MDS.U193 | 50K_Xba | 7 | 99,180,602 | qter | 59 |
| MDS.U29 | 50K_Xba | 7 | 99,180,602 | qter | 59 |
| MDS.U201 | 50K_Hind | 7 | 113,772,604 | qter | 45 |
| MDS.U311 | 250K_Nsp | 7 | 128,252,690 | qter | 30 |
| MDS.U14 | 50K_Xba | 8 | pter | qter | 144 |
| MDS.U344 | 250K_Nsp | 8 | 47,535,393 | qter | 99 |
| MDS.U61 | 50K_Xba | 9 | pter | 28,395,766 | 28 |
| MDS.U117 | 50K_Hind | 9 | pter | 36,941,035 | 37 |
| MDS.U128 | 50K_Hind | 10 | 73,570,575 | qter | 62 |
| MDS.U51 | 50K_Xba | 11 | pter | 48,202,912 | 46 |
| MDS.U143 | 250K_Nsp | 11 | pter | 57,701,308 | 57 |
| MDS.U308 | 250K_Nsp | 11 | pter | 47,692,599 | 47 |
| MDS.U2 | 50K_Xba | 11 | 3,945,802 | 35,547,219 | 2 |
| MDS.U108 | 50K_Hind | 11 | 43,508,116 | qter | 91 |
| MDS.U77 | 250K_Nsp | 11 | 48,191,389 | qter | 86 |
| MDS.U314 | 250K_Nsp | 11 | 56,227,770 | qter | 78 |
| MDS.U327 | 250K_Nsp | 11 | 57,603,286 | qter | 77 |
| MDS.U130 | 50K_Hind | 11 | 57,871,560 | qter | 76 |
| MDS.U161 | 50K_Hind | 11 | 61,780,852 | qter | 72 |
| MDS.U341 | 250K_Nsp | 11 | 63,631,519 | qter | 70 |
| MDS.U118 | 50K_Hind | 11 | 63,945,365 | qter | 70 |
| MDS.U19 | 50K_Xba | 11 | 65,679,725 | qter | 68 |
| MDS.U20 | 50K_Xba | 11 | 65,679,725 | qter | 68 |
| MDS.U309 | 250K_Nsp | 11 | 65,684,183 | qter | 68 |
| MDS.U306 | 250K_Nsp | 11 | 67,220,113 | qter | 67 |
| MDS.U307 | 250K_Nsp | 11 | 68,190,757 | qter | 66 |
| MDS.U325 | 250K_Nsp | 11 | 69,108,364 | qter | 65 |
| MDS.U70 | 50K_Xba | 11 | 70,697,060 | qter | 63 |
| MDS.U320 | 250K_Nsp | 11 | 73,254,396 | qter | 61 |
| MDS.U35 | 50K_Xba | 11 | 116,801,502 | 119,411,523 | 3 |
| MDS.U14 | 50K_Xba | 12 | 33,105,304 | qter | 99 |
| MDS.U309 | 250K_Nsp | 12 | 107,397,407 | qter | 24 |
| MDS.U192 | 50K_Hind | 13 | pter | qter | 95 |

| | | | | | |
|---|---|---|---|---|---|
| MDS.U322 | 250K_Nsp | 13 | pter | qter | 95 |
| MDS.U189 | 50K_Hind | 14 | pter | qter | 87 |
| MDS.U321 | 250K_Nsp | 14 | pter | qter | 87 |
| MDS.U71 | 50K_Xba | 14 | 34,289,096 | qter | 72 |
| MDS.U144 | 250K_Nsp | 14 | 36,420,563 | qter | 70 |
| MDS.U153 | 50K_Hind | 14 | 62,434,348 | qter | 44 |
| MDS.U141 | 250K_Nsp | 14 | 90,732,585 | qter | 16 |
| MDS.U141 | 250K_Nsp | 17 | pter | 16,322,726 | 16 |
| MDS.U103 | 50K_Hind | 17 | pter | 16,450,642 | 16 |
| MDS.U140 | 50K_Hind | 17 | pter | 20,055,907 | 20 |
| MDS.U66 | 50K_Xba | 17 | pter | 20,147,237 | 20 |
| MDS.U40 | 50K_Xba | 17 | pter | 22,048,100 | 22 |
| MDS.U58 | 50K_Xba | 17 | pter | 22,048,100 | 22 |
| MDS.U196 | 50K_Hind | 17 | pter | 27,217,068 | 27 |
| MDS.U102 | 50K_Hind | 17 | pter | 28,263,758 | 28 |
| MDS.U169 | 50K_Xba | 17 | 5,265,560 | 6,907,775 | 2 |
| MDS.U169 | 50K_Xba | 17 | 9,053,828 | 12,136,005 | 3 |
| MDS.U169 | 50K_Xba | 17 | 14,463,676 | 15,037,881 | 1 |
| MDS.U169 | 50K_Xba | 17 | 15,741,139 | 17,918,080 | 2 |
| MDS.U169 | 50K_Xba | 17 | 20,865,212 | 24,028,830 | 3 |
| MDS.U139 | 250K_Nsp | 17 | 23,046,031 | qter | 56 |
| MDS.U305 | 250K_Nsp | 17 | 25,755,541 | qter | 53 |
| MDS.U316 | 250K_Nsp | 17 | 37,251,667 | qter | 4 |
| MDS.U14 | 50K_Xba | 18 | pter | 42,628,679 | 42 |
| MDS.U161 | 50K_Hind | 18 | 14,366,731 | qter | 62 |
| MDS.U14 | 50K_Xba | 18 | 56,478,213 | qter | 19 |
| MDS.U43 | 50K_Xba | 19 | pter | 22,119,329 | 22 |
| MDS.U1 | 50K_Xba | 19 | 33,306,533 | qter | 30 |
| MDS.U325 | 250K_Nsp | 19 | 47,636,805 | qter | 16 |
| MDS.U251 | 250K_Nsp | 21 | pter | qter | 32 |
| MDS.U180 | 50K_Hind | 21 | 16,553,023 | qter | 30 |
| MDS.U130 | 50K_Hind | 21 | 17,968,550 | qter | 29 |
| MDS.U36 | 50K_Xba | 21 | 22,293,822 | qter | 25 |
| MDS.U177 | 50K_Hind | 22 | pter | qter | 34 |

**Supplementary Table 6. Summary of 11q-aUPD and *c–Cbl* mutations**

| Sample ID | Age | Sex | WHO Classification | IPSS† | Copy number changes in CNAG | 11qUPD | *c–Cbl* mutations | | Origin of c-Cbl mutation | Coexisting gene mutations§ |
|---|---|---|---|---|---|---|---|---|---|---|
| MDS.U77 | 76 | M | CMML1 | Low | None | + | 1238A>C<br>1349G>A | Q367P¦<br>C404Y¦ | Somatic<br>Somatic | |
| MDS.U130 | 47 | M | CMML2 | HIGH | None | + | 1288T>G | C384G | ND¶ | *RUNX1*/R161K |
| MDS.U118 | 89 | F | CMML1 | Low | None | + | 1237C>A | Q367K | ND | *RUNX1*/F330L |
| MDS.U161 | 35 | M | CMML1 | INT–1 | None | + | 1334T>C | L399P | ND | |
| MDS.U19 | 68 | M | CMML2 | INT–2 | None | + | 1285A>C<br>1238A>C | I383L&<br>Q367P& | ND | *JAK2/V617F* |
| MDS.U70 | 80 | M | CMML2 | HIGH | None | + | 1250A>G | Y371C | ND | |
| MDS.U306 | 30 | M | CMML2 | HIGH | None | + | 1286T>G | I383R | ND | *RUNX1*/Q293X |
| MDS.U307 | 59 | M | CMML1 | LOW/INT-1 | None | + | 1390T>G | F418V | Somatic | |
| MDS.U314 | 82 | F | CMML1 | LOW/INT-1 | None | + | 1325G>A | C396Y | ND | *JAK2/V617F* |
| MDS.U320 | 73 | F | CMML1 | INT-1/INT-2 | None | + | △1343-1351 | △369-371 | ND | *JAK2/V617F* |
| MDS.U325 | 67 | F | CMML1 | LOW | None | + | 1250A>C | Y371S | Somatic | |
| MDS.U341 | 56 | M | CMML2 | INT-2 | None | + | 1331A>G | H398R | ND | *RUNX1*/Ala115fsX458 |
| MDS.U309 | 74 | F | CMML2 | INT-2/HIGH | i17q | + | 1384T>Cå | C416R | ND | |
| MDS.U35 | 64 | M | RAEB1 | INT–2 | del(5)(q22.1),del(5)(q31.1–33.2),del(7)(p22.3–11.2),del(7)(q11.23–21.11),add(7)(q21.11–21.2),del(7)(q22.1–34),add(8)(q13.1–24.3),del(9)(p24.3–23),del(9)(p22.2–21.3),del(9)(p21.3–21.2),del(9)(p21.1–q21.32),del(9)(q22.2–31.2),del(11)(q22.2–23.3),del(11)(q23.3–25),del(12)(p13.2–11.21),del(12)(q12) | + | 1250A>C | Y371S | ND | *p53*/I251N |
| MDS.U108 | ND | F | AML# | NA | del(5)(q23.2–32),–7,del(12)(p13.2) | + | ΔIntron7/Exon8‡ | △368-382 | ND | |
| MDS.U327 | 41 | M | CMML2 | INT-1/INT-2 | +8, +21 | + | – | – | ND | *RUNX1*/Trp79X |
| MDS.U20 | 68 | M | RAEB2 | INT-2 | None | + | – | – | NA | |
| MDS.U342 | 77 | F | CMML1 | LOW | None | – | 1289G>A<br>1165C>T | C384Y‖<br>R343X‖ | ND | |
| MDS.U337 | 60 | M | CMML1 | INT-1 | del(11)(q22.3),del(12)(p13.31-p13.1) | – | 1396C>G | R420G | ND | *FLT3 ITD*? |
| MDS.U189 | 64 | M | AML# | NA | None | – | 1241G>A | Y368C | ND | |

NA: not applicable

† The International Prognostic Scoring System; NA: not applicable

‡ 212bp deletion in boundary between intron 7 and exon 8, corresponding to amino acids 367-382 (Δ367-382 in Fig. 1d)

¶ Not determined, because no normal tissue samples were available

§ Coexisting *Runx1*-involvong translocations with *c-Cbl* mutations were also reported in AML (Abbas et al., 2008)

# AML with myelodysplasia-related changes

¦ Two mutations occurred on the separate alleles, most likely in different clones

& Q367P mutation occurred on the pre-existing I383L allele in a minor subclone

‖ C384Y was the predominant allele, but phase determination was unsuccessful

**Supplementary Table 7. Characteristics of myeloid neoplasms either with or without *c–Cbl* mutations**

| | All cases | | Cases with normal karyotypes[£] | | Cases with CMML | |
|---|---|---|---|---|---|---|
| **Mutation status of *c–Cbl*** | **Mut (+)** | **Mut (–)** | **Mut (+)** | **Mut (–)** | **Mut (+)** | **Mut (–)** |
| Number of cases | 9 | 162 | 7 | 63 | 6 | 29 |
| Median age (y.o.)[§] | 69 | 67 | 68 | 67 | 68 | 70 |
| Sex (M/F) | 6/3 | 113/49 | 6/1 | 44/19 | 6/0 | 10/19 |
| UPD in 11q | | | | | | |
| (+) | 8 | 1 | 7 | 1 | 6 | 0 |
| (–) | 1 | 161 | 0 | 62 | 0 | 29 |
| Karyotypes | | | | | | |
| Normal | 7 | 63 | 7 | 63 | 6 | 12 |
| Complex[$] | 1 | 22 | 0 | 0 | 0 | 0 |
| aUPD | 9 | 42 | 7 | 13 | 6 | 10 |
| FAB classification[†] | | | | | | |
| CMML | 6 | 29 | 6 | 13 | 6 | 29 |
| RAEBt/MDS–AML | 2 | 37 | 1 | 4 | 0 | 0 |
| RAEB | 1 | 57 | 0 | 21 | 0 | 0 |
| RA/RARS | 0 | 34 | 0 | 11 | 0 | 0 |
| WHO classification | | | | | | |
| MDS | | | | | | |
| RCUD/RARS | 0 | 4 | 0 | 2 | 0 | 0 |
| RCMD | 0 | 25 | 0 | 11 | 0 | 0 |
| MDS associated with isolated del(5q) | 0 | 6 | 0 | 0 | 0 | 0 |
| RAEB–1 | 1 | 27 | 1 | 13 | 0 | 0 |
| RAEB–2 | 0 | 42 | 0 | 16 | 0 | 0 |
| MDSu | 0 | 4 | 0 | 3 | 0 | 0 |
| MDS/MPD | | | | | | |
| CMML–1 | 3 | 16 | 3 | 7 | 2 | 16 |
| CMML–2 | 3 | 13 | 3 | 7 | 3 | 13 |
| AML | | | | | | |
| AML with myelodysplasia-related changes[¶] | 2 | 22 | 2 | 4 | 0 | 0 |
| Therapy-related myeloid neoplasms | 0 | 4 | 0 | 0 | 0 | 0 |
| IPSS[¶] | | | | | | |
| HIGH | 2 | 33 | 2 | 5 | 2 | 5 |
| INT–2 | 2 | 39 | 1 | 17 | 1 | 9 |
| INT–1 | 1 | 53 | 1 | 29 | 1 | 11 |
| LOW | 2 | 10 | 2 | 5 | 2 | 2 |
| Median survival (m)[§] | 10.5 | 13.6 | 18.9 | 17.8 | 10.5 | 15 |
| Transformation to AML[§] | 6/8 (75.0%) | 88/137 (64.2%) | 4/6 (66.7%) | 31/55 (56.4%) | 3/5 (60.0%) | 16/24 (66.7%) |

[†] The French–American–British Classification
[‡] The World Health Organization Classification
[¶] The International Prognostic Scoring System for informative cases
[$] Defined as karyotypes having more than 4 abnormalities.
[£] Normal copy number profile in SNP array analysis
[§] Calculations for informative cases

**Supplementary Table 8. Primer sets and their PCR conditions used in this study**

| *Gene* | Exon | F/R[†] | Sequence[‡] | Annealing temperature | DMSO[¶] | Betain[§] |
|---|---|---|---|---|---|---|
| *RUNX1* | 3 | F | 5'–AGCTGTTTGCAGGGTCCTAA–3' | 56˚C | | |
| *RUNX1* | 3 | R | 5'–GTCCTCCCACCACCCTCT–3' | | | |
| *RUNX1* | 4 | F | 5'–CATTGCTATTCCTCTGCAACC–3' | 60˚C | | |
| *RUNX1* | 4 | R | 5'–CCATGAAACGTGTTTCAAGC–3' | | | |
| *RUNX1* | 5 | F | 5'–CCACCAACCTCATTCTGTTT–3' | 56˚C | | |
| *RUNX1* | 5 | R | 5'–AGACATGGTCCCTGAGTATA–3' | | | |
| *RUNX1* | 6 | F | 5'–AGCCCCAGTTTTAGGAAATCCAC–3' | 60˚C | | 1M |
| *RUNX1* | 6 | R | 5'–GAGCATCAAGGGGAAACCCC–3' | | | |
| *RUNX1* | 7 | F | 5'–CCCACCCCACTTTACATATAATTG–3' | 56˚C | | |
| *RUNX1* | 7 | R | 5'–CCAGCTCAGCTGCAAAGAATGTG–3' | | | |
| *RUNX1* | 8 | F | 5'–CGCAACCTCCTACTCACTTCCG–3' | 60˚C | | |
| *RUNX1* | 8 | R | 5'–GGGCTTGTCGCGAACAGGAGG–3' | | | |
| | | | | | | |
| *N–Ras* | 2 | F | 5'–AAGCTTTAAAGTACTGTAGA–3' | 56˚C | | |
| *N–Ras* | 2 | R | 5'–AAGCTTATTGCATAACTGAA–3' | | | |
| *N–Ras* | 3 | F | 5'–GATTCTTACAGAAAACAAGT–3' | 56˚C | 5% | 1M |
| *N–Ras* | 3 | R | 5'–GTAGAGGTTAATATCCGCAA–3' | | | |
| | | | | | | |
| *TP53* | 5 | F | 5'–TTCCTCTTCCTGCAGTACTC–3' | 56˚C | | |
| *TP53* | 5 | R | 5'–GCAAATTTCCTTCCACTCGG–3' | | | |
| *TP53* | 6 | F | 5'–ACCATGAGCGCTGCTCAGAT–3' | 56˚C | | |
| *TP53* | 6 | R | 5'–AGTTGCAAACCAGACCTCAG–3' | | | |
| *TP53* | 7 | F | 5'–GTGTTGTCTCCTAGGTTGGC–3' | 62˚C | | |
| *TP53* | 7 | R | 5'–CAAGTGGCTCCTGACCTGGA–3' | | | |
| *TP53* | 8 | F | 5'–CCTATCCTGAGTAGTGGTAA–3' | 56˚C | | |
| *TP53* | 8 | R | 5'–CCAAGACTTAGTACCTGAAG–3' | | | |
| | | | | | | |
| *FLT3* | ITD | F | 5'–GCAATTTAGGTATGAAAGCCAGC–3' | 62˚C | | |
| *FLT3* | ITD | R | 5'–CTTTCAGCATTTTGACGGCAACC–3' | | | |
| | | | | | | |
| *MPL* | 1 | F | 5'–GGAGGATGGGCTAAGGCAG–3' | 60˚C | | 1M |
| *MPL* | 1 | R | 5'–TCTTCCTGGGGCATAGGTGA–3' | | | |
| *MPL* | 2 | F | 5'–CCCTTCCACATAAACATGCCT–3' | 62˚C | | |
| *MPL* | 2 | R | 5'–GCAGGAAAGCTGCTGGAGT–3' | | | |
| *MPL* | 3 | F | 5'–GTCCTCAGGCGTCCGCAT–3' | 62˚C | | 1M |
| *MPL* | 3 | R | 5'–GGTATCCGTGCTGAGCTGGA–3' | | | |
| *MPL* | 4 | F | 5'–TCCAGAGGCTGAGCCATAGAC–3' | 62˚C | | |
| *MPL* | 4 | R | 5'–GGTCTGGAATCCCCAAAGT–3' | | | |
| *MPL* | 5 | F | 5'–GGTTGGAGGCTCTCTCAGCT–3' | 62˚C | | |
| *MPL* | 5 | R | 5'–CTTTTATCTCCTCCCCATCTCC–3' | | | |
| *MPL* | 6 | F | 5'–CCTATACAGTAGGGGCACACG–3' | 60˚C | | |
| *MPL* | 6 | R | 5'–TGTGGCTCACTCCCATGACA–3' | | | |
| *MPL* | 7 | F | 5'–GATGGGAAGCCTTGGGATTAG–3' | 60˚C | | |
| *MPL* | 7 | R | 5'–GGGAACTATGTGGAAGAAT–3' | | | |
| *MPL* | 8 | F | 5'–CCTTGTGCACAGAAGGACTTA–3' | 62˚C | | |
| *MPL* | 8 | R | 5'–CCCCTGCGTAGTGAGGTCTG–3' | | | |
| *MPL* | 9 | F | 5'–CGAAGCCCCGACGCCGGGCCA–3' | 62˚C | 5% | 1M |
| *MPL* | 9 | R | 5'–CAGGCGCTGTGCGGCTTTGG–3' | | | |
| *MPL* | 10 | F | 5'–AGGGGCGGGGCCAGAGTA–3' | 62˚C | | 1M |
| *MPL* | 10 | R | 5'–AGAGGTGACGTGCAGGAA–3' | | | |
| *MPL* | 11 | F | 5'–CTGCCAATCCACTGCCATG–3' | 62˚C | | 1M |
| *MPL* | 11 | R | 5'–AGTACCAGGCAGGGTTGGTG–3' | | | |
| *MPL* | 12 | F | 5'–TCCCACAGGATCTGCTTTAAT–3' | 62˚C | | 1M |
| *MPL* | 12 | R | 5'–GAGTTTAGCTCTGTCCAGGGAAC–3' | | | |
| | | | | | | |
| *JAK2* | 14 | F | 5'–TGCTGAAAGTAGGAGAAAGTGC–3' | 60˚C | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| *JAK2* | 14 | R | 5'–CTGACACCTAGCTGTGATCCTG–3' | | | |
| *c–Cbl* | 1 | F | 5'–CTTCACGCCCTGCTTCTCT–3' | 64˚C | 5% | 1M |
| *c–Cbl* | 1 | R | 5'–CCAGCACTTCTCCACCATCT–3' | | | |
| *c–Cbl* | 2 | F | 5'–TGGGGTTATGGATCTGCTTC–3' | 60˚C | | |
| *c–Cbl* | 2 | R | 5'–TCAAATGTCAACTTTCCATTC–3' | | | |
| *c–Cbl* | 3 | F | 5'–CATCTTGTATGGTGAATTTGGTG–3' | 60˚C | | |
| *c–Cbl* | 3 | R | 5'–AACCAAAGCCAGGAAATACATA–3' | | | |
| *c–Cbl* | 4 | F | 5'–GTGGCTCTCCTTCCTTTCCT–3' | 60˚C | | |
| *c–Cbl* | 4 | R | 5'–CAGACAAAAACCCCAAATGTT–3' | | | |
| *c–Cbl* | 5 | F | 5'–ATTGCCCTCTGAGTTGGTTG–3' | 60˚C | | |
| *c–Cbl* | 5 | R | 5'–AACCTTGGCTATTGCGAAAC–3' | | | |
| *c–Cbl* | 6 | F | 5'–GCTGCAGTGAGCTGAGATTG–3' | 60˚C | | 1M |
| *c–Cbl* | 6 | R | 5'–TCCAGGACCTGAAGGTCAAA–3' | | | |
| *c–Cbl* | 7 | F | 5'–CTTACACCACGTTGCCCTTT–3' | 60˚C | | |
| *c–Cbl* | 7 | R | 5'–TGGGTCCTATTTTAAGCTCCA–3' | | | |
| *c–Cbl* | 8 | F | 5'–AGGACCCAGACTAGATGCTTTC–3' | 60˚C | | |
| *c–Cbl* | 8 | R | 5'–GGCCACCCCTTGTATCAGTA–3' | | | |
| *c–Cbl* | 9 | F | 5'–CTGGCTTTTGGGGTTAGGTT–3' | 60˚C | | |
| *c–Cbl* | 9 | R | 5'–TCGTTAAGTGTTTTACGGCTTT–3' | | | |
| *c–Cbl* | 10 | F | 5'–AACCTAGGTCTGGCCCATTT–3' | 60˚C | | |
| *c–Cbl* | 10 | R | 5'–GTGTGGGAGGTACTGGCAAT–3' | | | |
| *c–Cbl* | 11 | F | 5'–TGCAGTTTTTGGATTCTTTGC–3' | 62˚C | 5% | 1M |
| *c–Cbl* | 11 | R | 5'–ACAGACATGAGCCACTGTGC–3' | | | |
| *c–Cbl* | 12 | F | 5'–TTCATTTCCCTGGCTCTTTC–3' | 60˚C | | |
| *c–Cbl* | 12 | R | 5'–TGGGTTTTCTCAATTTTCTGC–3' | | | |
| *c–Cbl* | 13 | F | 5'–AAGCTGCACCTTTGGTCAGT–3' | 60˚C | | |
| *c–Cbl* | 13 | R | 5'–AGCAGGTGAAGGGTGTCAAT–3' | | | |
| *c–Cbl* | 14 | F | 5'–GCAAAACGAGAAGATGAA–3' | 60˚C | | 1M |
| *c–Cbl* | 14 | R | 5'–AGCCTGGCCAACATAGTGAA–3' | | | |
| *c–Cbl* | 15 | F | 5'–CTGCAGTGTTGGTGCAGATT–3' | 60˚C | | |
| *c–Cbl* | 15 | R | 5'–CTTGGGACTTTCCTCCCATT–3' | | | |
| *c–Cbl* | 16 | F | 5'–AAAACCCAGCCTTGTGACTG–3' | 60˚C | | |
| *c–Cbl* | 16 | R | 5'–GGAACTCCTGCCCTTCTAGG–3' | | | |

† F: Forward, R: Reverse

‡ Used for both PCR amplification and sequencing

¶ Final concentration (v%)

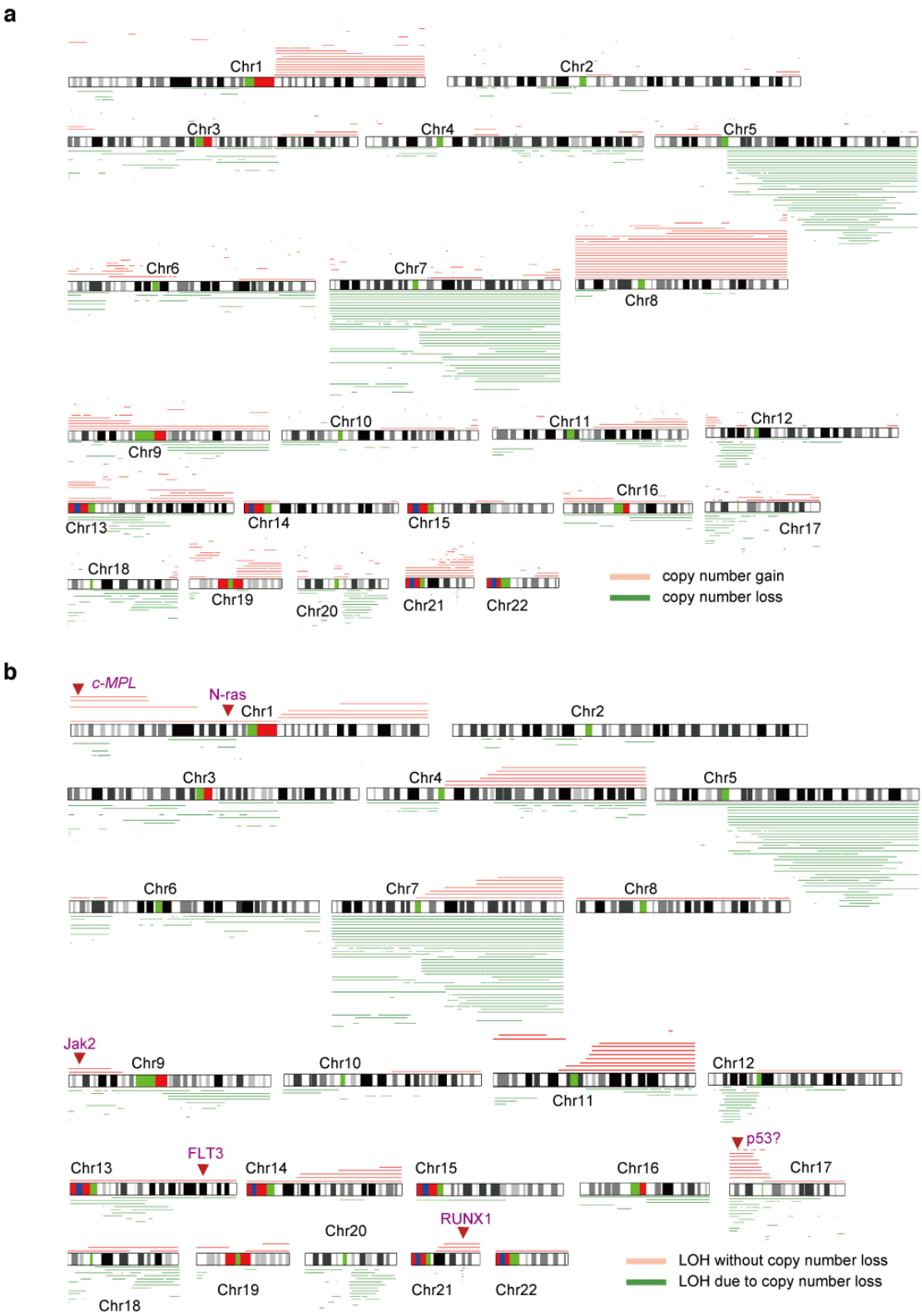
§ Final concentration (mol/L)

**Supplementary Table 9. Antibodies used in this study**

| Antigen | Species[†] | Purpose[‡] | Conjugate | ID | Supplier |
|---|---|---|---|---|---|
| Phosphotyrosine | H/M | WB | NA[§] | 05–321(4G10) | Upstate |
| Akt | H/M | WB | NA | #9272 | Cell–Signaling |
| c–Cbl | H/M | WB | NA | sc–170 | Santa Cruz |
| Cbl–b | H/M | WB | NA | sc–8006 | Santa Cruz |
| EGFR | H/M | IP | NA | sc120 | Santa Cruz |
| EGFR | H/M | WB | NA | 06–129 | Upstate |
| HA.11 | NA | WB/IP | NA | MMS–101R | COVANCE |
| HA.11 | NA | WB/IP | NA | PRB–101C | COVANCE |
| Phospho–Akt(Ser473) | H/M | WB | NA | #9271 | Cell–Signaling |
| Ubiquitin | H/M | WB | NA | sc–8017 | Santa Cruz |
| Stat5 | H/M | WB | NA | #9310 | Cell–Signaling |
| Phospho-Stat5(Tyr694) | H/M | WB | NA | #9351 | Cell–Signaling |
| Flt3 | H/M | WB/IP | NA | sc-479 | Santa Cruz |
| JAK2 | H/M | WB/IP | NA | sc-278 | Santa Cruz |
| c-kit | H/M | WB/IP | NA | sc-168 | Santa Cruz |
| phospho-c-kit(Tyr719) | H/M | WB | NA | PK1011 | Calbiochem |
| FLAG | NA | WB/IP | NA | F7425 | SIGMA |
| FLAG | NA | WB/IP | NA | F1804 | SIGMA |
| CD4 | M | FACS | Biotin | | BD Pharmingen |
| CD8 | M | FACS | Biotin | | BD Pharmingen |
| B220 | M | FACS | Biotin | | BD Pharmingen |
| Gr–1 | M | FACS | Biotin | | BD Pharmingen |
| TER–119 | M | FACS | Biotin | | BD Pharmingen |
| CD34 | M | FACS | FITC | | BD Pharmingen |
| CD16/32 | M | FACS | PE | | BD Pharmingen |
| c–kit | M | FACS | APC | | BD Pharmingen |
| Sca–1 | M | FACS | PE/Cy7 | | BD Pharmingen |

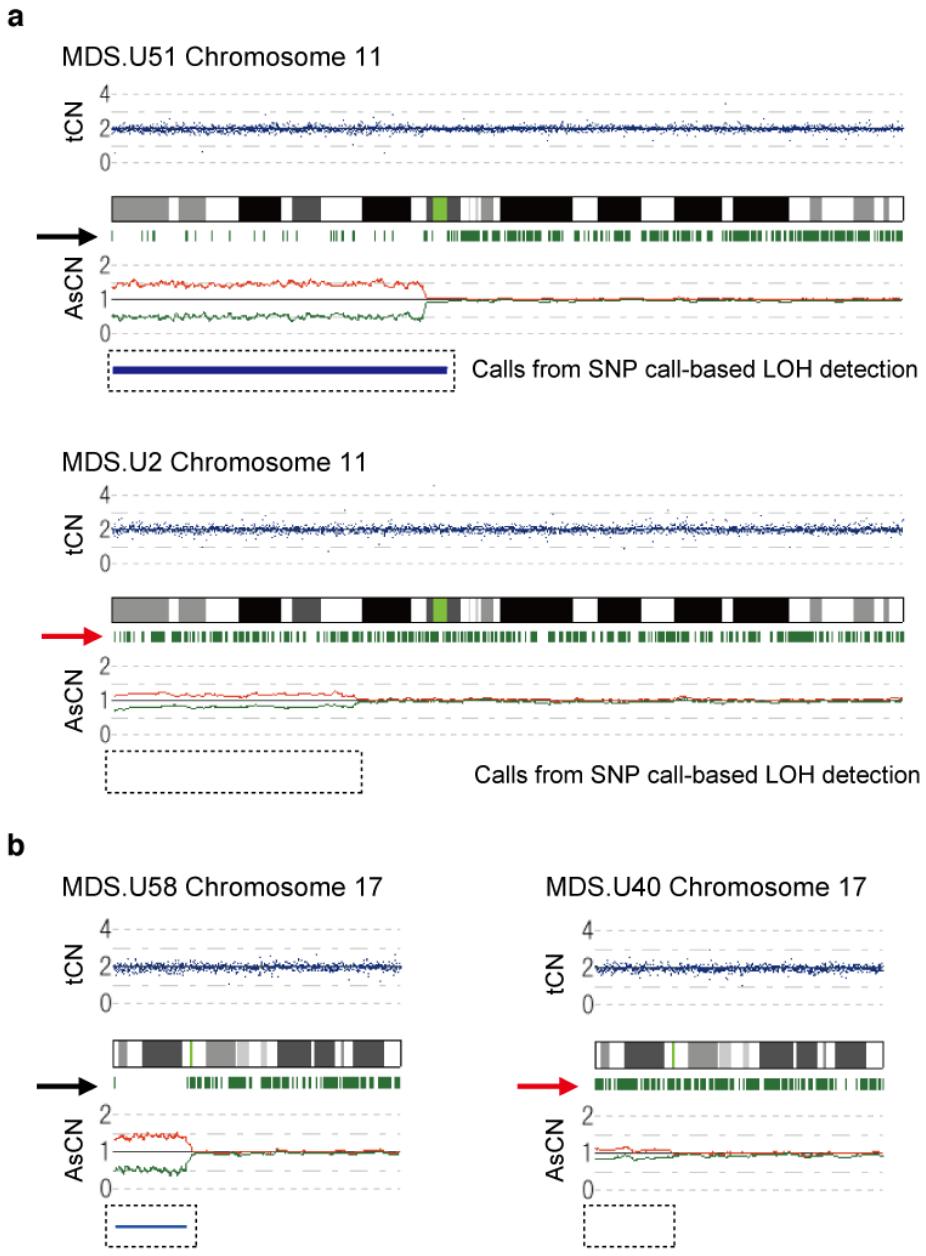
[†] H: Human, M: Mouse
[‡] WB: Western Blot, IP: Immunoprecipitation, FACS: Flow cytometric analysis and FACS
[§] NA: not applicable

## Supplementary Figure 1



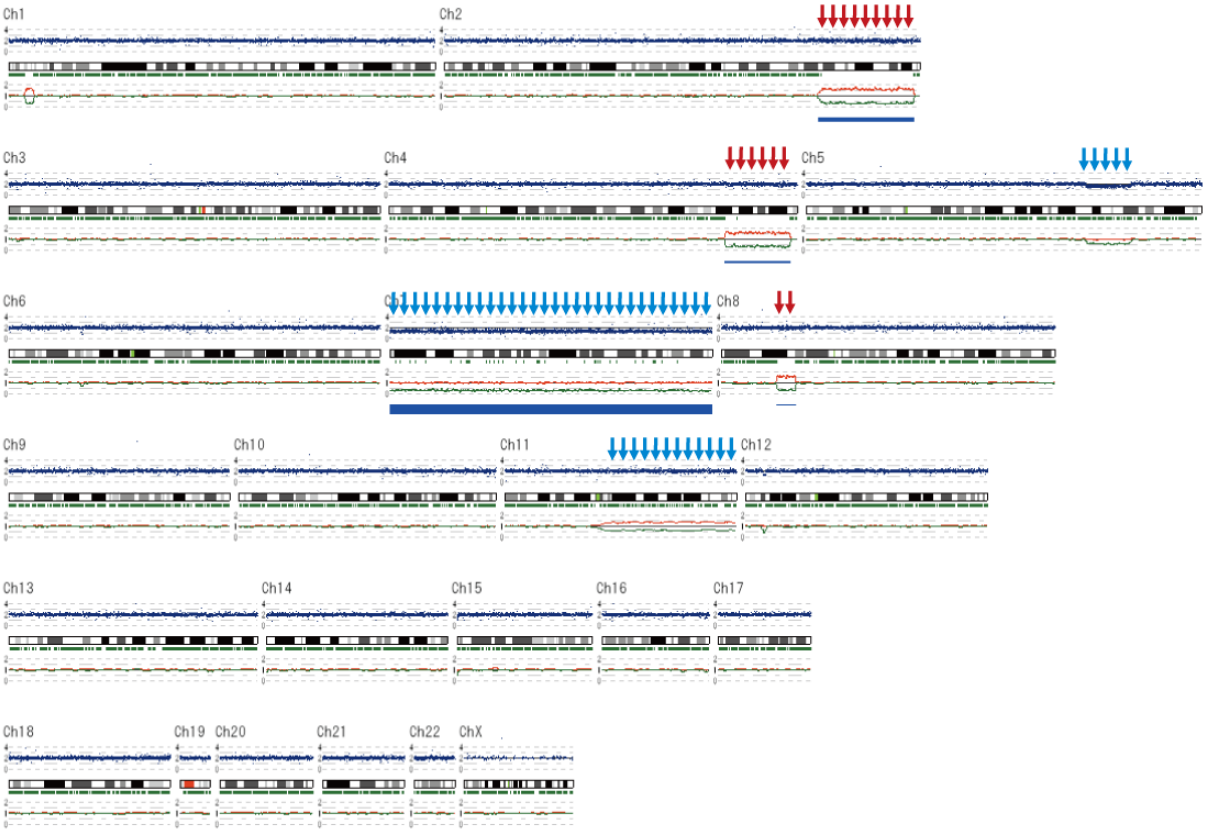

**Integrated pictures of copy alterations and LOH on ideograms in myeloid neoplasms.**
**a**, Copy number gains (red bars) and losses (green bars) found in 222 samples of myeloid neoplasms are depicted on ideogram of each chromosome. Each horizontal bar represents an abnormality found in a single sample. **b**, Distribution of LOH in the same sample set. LOHs with copy number loss are shown by green bars and copy number-neutral LOHs by red bars. Note that the distribution of LOH due to copy number loss is identical to that of copy number loss in panel **a**. Putative gene targets of UPD clusters in 1p (*c–MPL or N–ras*), 9p (*Jak2*), 13q (*FLT3*),17p (*p53*), and 21q (*Runx1*), are indicated by arrow heads.

## Supplementary Figure 2




**Sensitive detection of copy number neutral LOH (or aUPD) by AsCNAR**
Total copy numbers (tCN) are plotted in blue, and allele–specific copy numbers (AsCN) in red and green for larger and smaller alleles, respectively. Heterozygous SNP calls are shown in green bars below each ideogram. In SNP array analysis, the presence of LOH can be easily detected by significantly reduced heterozygous SNP calls in a given region compared with their expected frequencies, even when it does not accompany any copy number reduction (aUPD) (SNP call–based LOH detection, e.g., see ref 7). In MDS.U51 (a, upper panel) and MDS.U58 (**b**, left panel), 11p and 17p show normal tCN, but heterozygous SNP calls within these regions mostly disappear (black arrowheads), predicting the presence of copy number neutral LOH (horizontal blue bar), although discrimination of shared allele by their parents is difficult. However, there is no apparent reduction of heterozygous SNP calls found in the similar regions in MDS.U2 (a, lower panel) and MDS.U40 (b, right panel), and thus no LOH calls by SNP call–based LOH detection. Even in these cases, AsCNs in these regions are clearly dissociated (red arrowheads), sensing the presence of LOH in a minor tumor population. In the latter cases, these regions represent true aUPD, because of the presence of background heterozygous SNP calls. Primary tumor specimens frequently suffer from “contamination” to varying degrees with normal cell components. If the normal component exceeds 30% of the population, SNP call–based LOH detection fails (see ref 7). AsCNAR, however, has the sensitivity to detect these LOH regions, relying on subtle imbalances in allele–specific microarray signals.

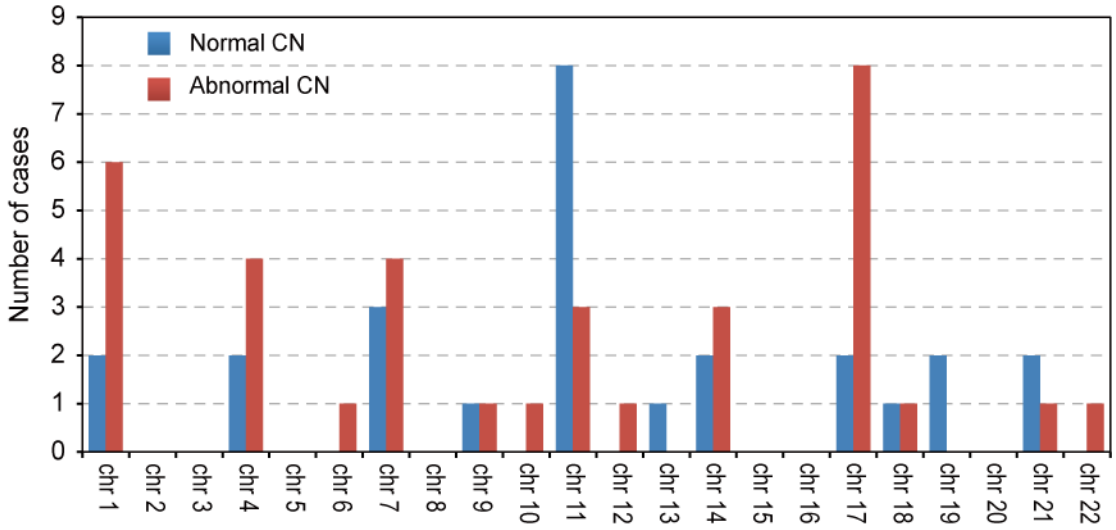
## Supplementary Figure 3




**Shared identical-by-descent (IBD) alleles by parents versus aUPD**

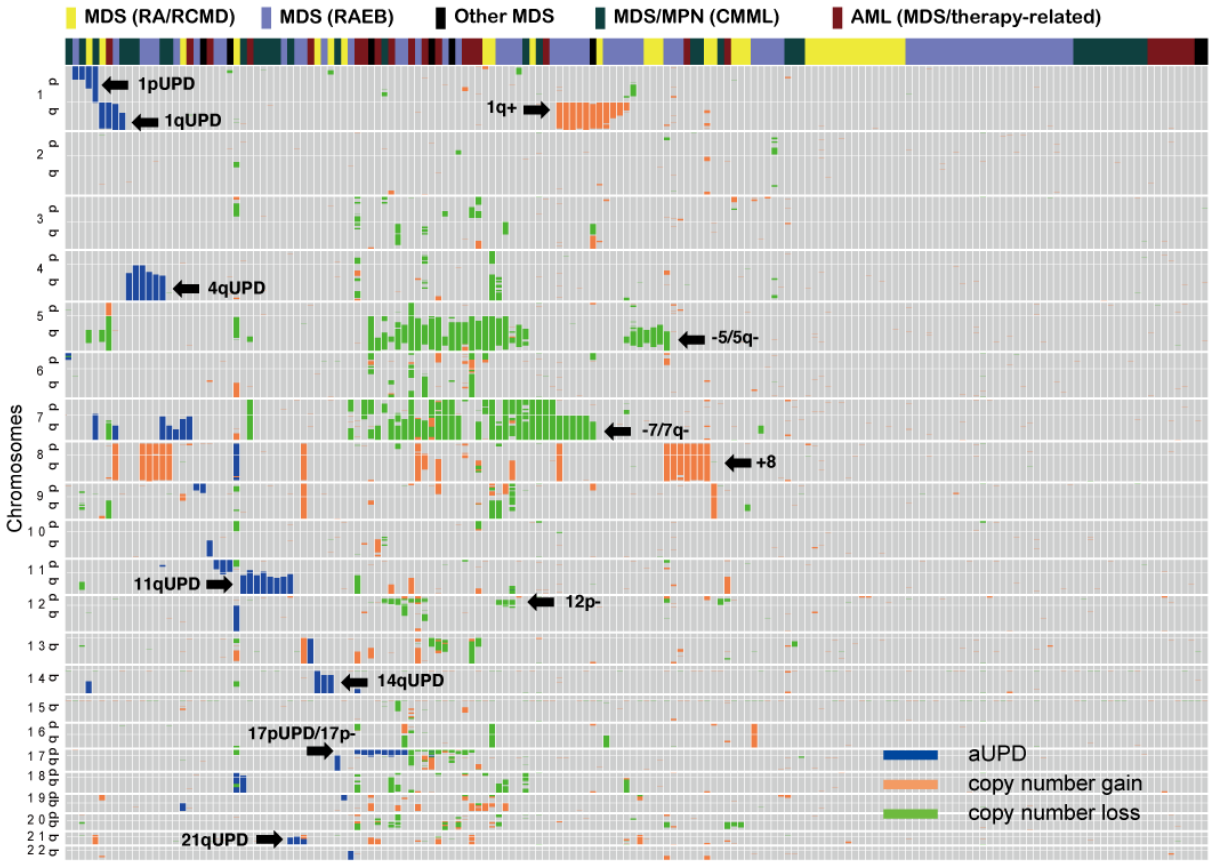
In SNP array analysis, aUPD appears as a long stretch of homozygous SNP calls, or significantly reduced heterozygous SNP calls as compared to the frequencies expected in general population. In the absence of constitutive DNA, as is frequently the case with analysis of MDS, researchers base detection of aUPD on this phenomenon. This leads to false positive detection of aUPD through inclusion of those genomic segments that mostly contains homozygous SNPs, which might be due to shared IBD alleles by one's parents, most likely due to some degrees of consanguinity that was more prevalent previously. In fact, previously nearly 10% of marriages were between close kin in Japan. Fortunately, according to our unpublished observation using paired DNA samples from many cancer patients, when such a long homozygous tract occurs within a diploid segment and its both ends are flanked by heterozygous diploid alleles (red arrows in chromosomes 2, 4, and 8), it represents alleles shared by parents, not by an aUPD (unpublished data). We exclude such homozygous stretches from the current study. However, an acquired UPD frequently accompanies some degrees of heterozygous SNP calls due to the presence of normal cell components. In some cases, the reduction in heterozygous SNP calls is not significantly reduced, which occurs when normal cell components exceed ~30%[10]. Even in this situation, AsCNAR detects the presence of an aUPD by sensing subtle distortions caused by allelic imbalances within the samples. Note that in the long arm of chromosome 11, reduction of heterozygous SNP calls was barely detected, but AsCNAR sensed the presence of allelic imbalance as identified by the presence of dissociation between allelic copy numbers. The degrees of reduction in heterozygous SNP calls may differ between segments, which implies the presence of heterogeneity of neoplastic clones. In these bone marrow specimens, a very high tumor content was expected because heterozygous SNP calls mostly disappeared in chromosome 7, indicating most tumor cells carried monosomy 7. However, reduction of heterozygous SNP calls was incomplete in chromosomes 5 and 11. Thus MDS clones having del(5q) and 11qUPD only account for sub–populations.

**Supplementary Figure 4**




**Chromosomal distribution of aUPD in cases with or without copy number abnormalities**
The number of aUPD in each chromosome is shown in cases with normal (blue bar) and abnormal (red bar) CN profiles.

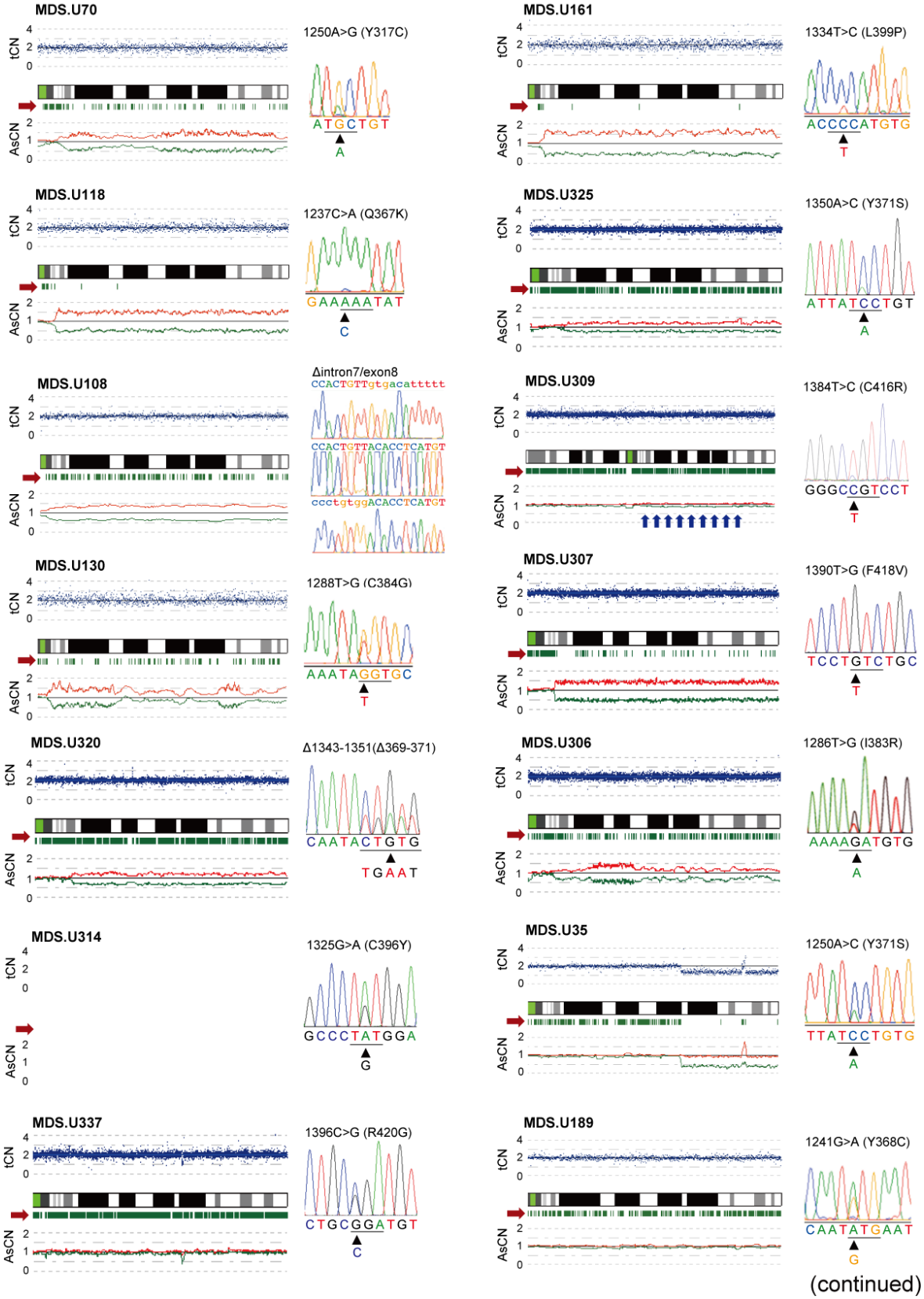
## Supplementary Figure 5



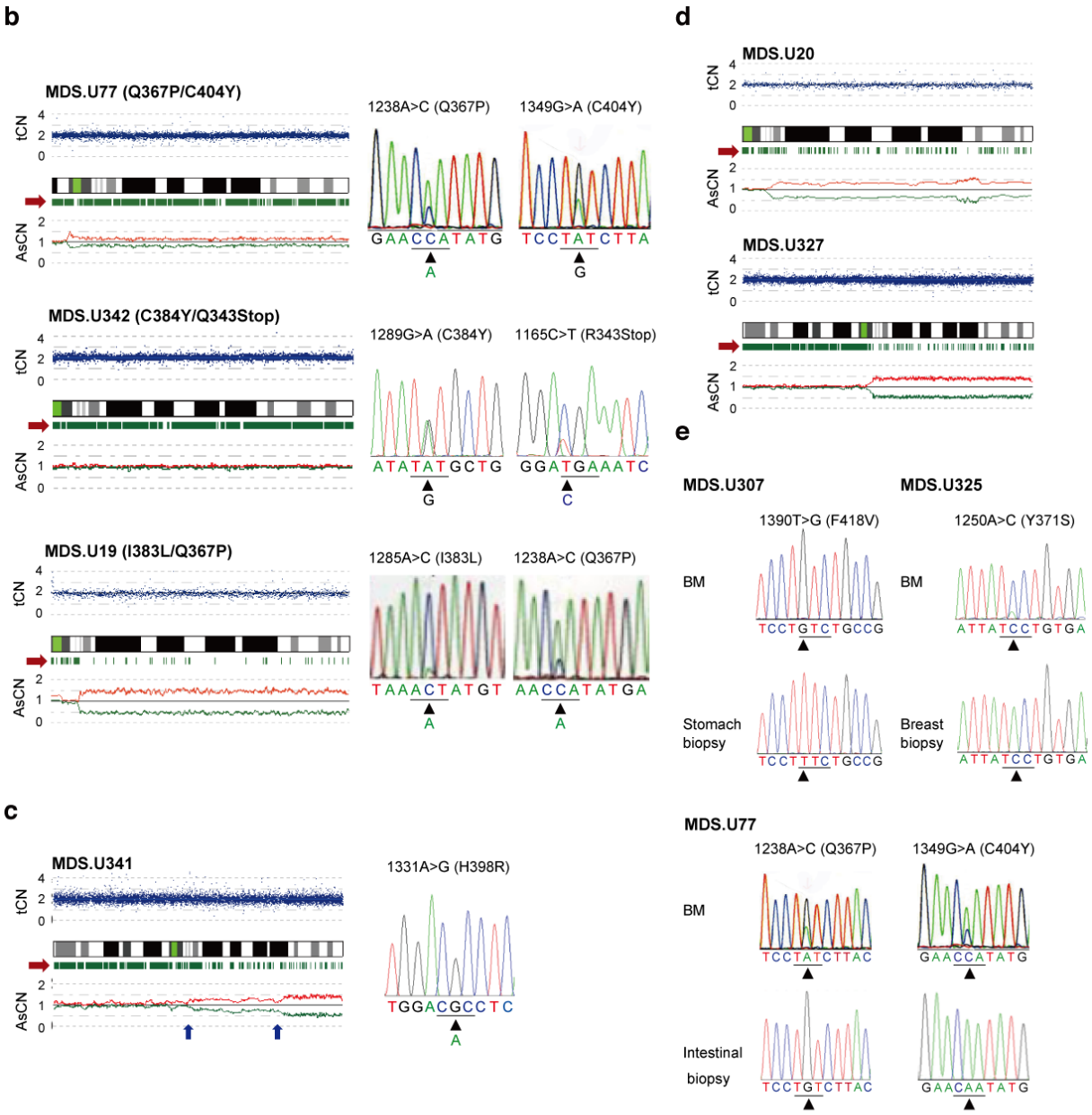

**Distributions of genetic lesions in 222 samples with myeloid neoplasms.**
Allele-specific copy numbers (CNs) were inferred from raw signal ratios at each SNP locus for each specimen using CNAG/AsCNAR software. The genetic lesions, including CN gains, losses, and aUPD, are color-coded and plotted vertically in chromosomal order for each sample; from 1p (top) to 22q (bottom), as indicated. Vertical position of each lesion are proportional to the genetic length and thus the size of the colored boxes corresponds to the length of DNA affected. Samples were horizontally clustered based on the Pearson's correlation coefficient with regard to genomic CNs and subsequently rearranged by visual inspections. Major histology types are indicated by different colors on the top panel as indicated. Note that some genetic changes involving small regions are lost in this figure due to limited resolution. aUPD in particular chromosomal arms tends to be found in mutually exclusive cases, enabling clustering based on the site of aUPD, except for 17qLOH, which was frequently accompanied by 5q-/-5, 7q-/-7, and 12p-. See also Supplementary Figures 1 and 2. Clusters of common genetic lesions are indicated by arrows.

## Supplementary Figure 6

**a**



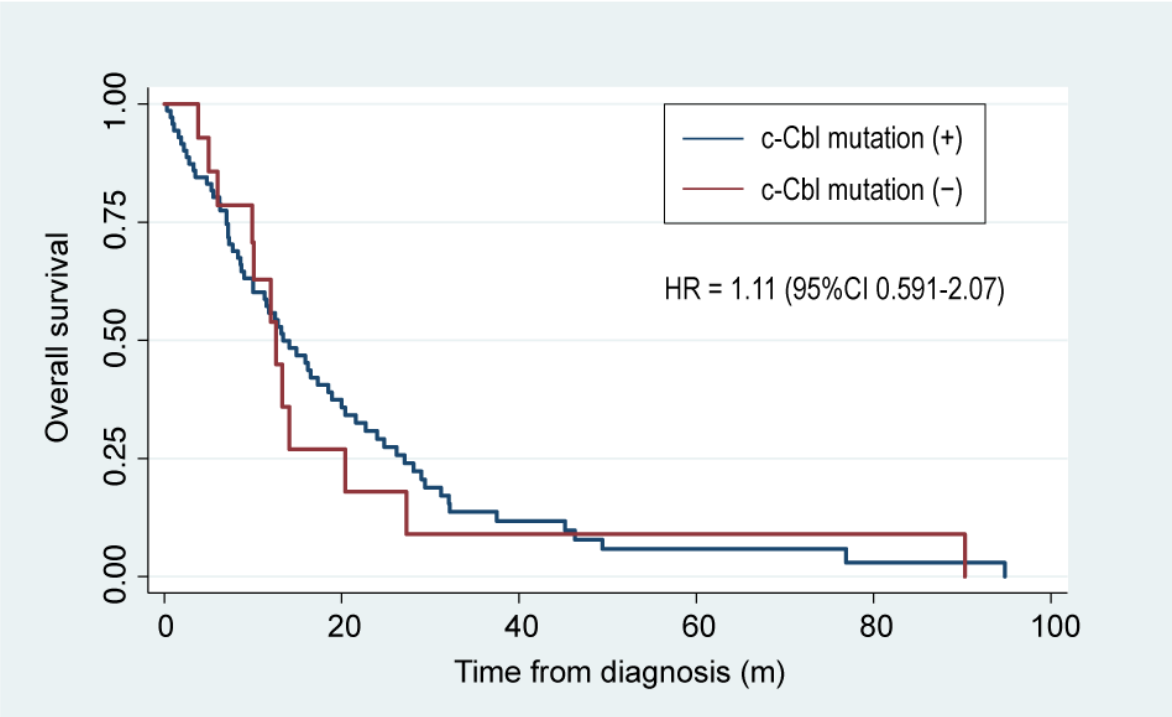

(continued)




**aUPD in 11q arm and/or *c–Cbl* mutations in myeloid neoplasms**

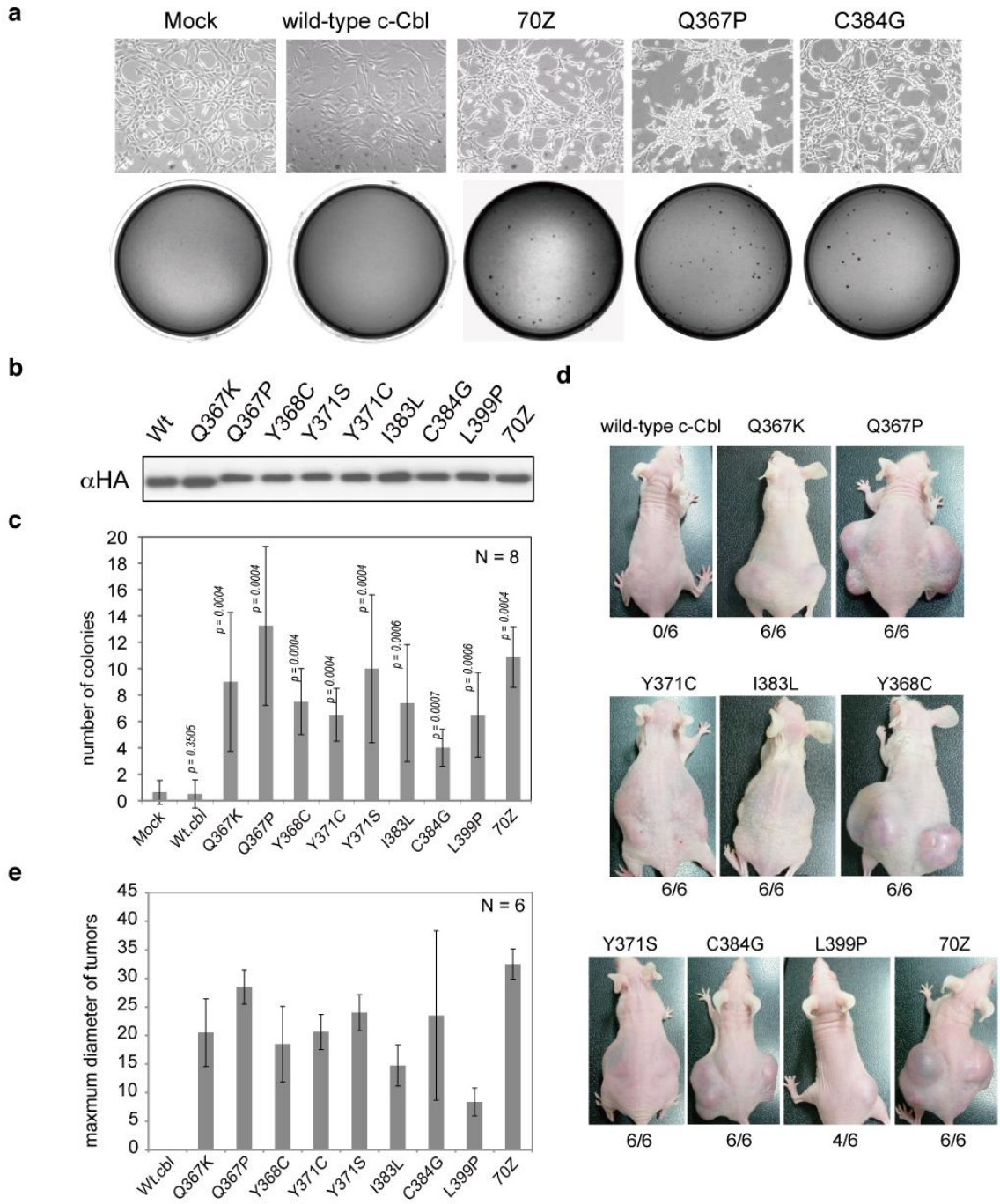
**a-c**, CNAG outputs for total copy number (tCNs) and allele–specific copy number (AsCNs) in the long arm of chromosome 11 in 17 samples with c-Cbl mutations with or without 11q-aUPD, including 3 cases harboring dual mutations (**b**). Chromatograms showing *c-Cbl* mutations are presented in the right. The green bars below each ideogram of 11q (red arrow) indicate the positions of heterozygous SNP calls. Substantial heterozygous SNP calls in the presence of split AsCNs suggest that the dissociated AsCNs represent aUPD but not shared IBD alleles. AsCNs are plotted after moving averages of 10 adjacent SNPs were taken. Dissociation of AsCN plots indicates the presence of UPD. Proportion of the UPD component is evaluable from the size of the dissociation in AsCNs. MDS.U341 showed two discrete breakpoints in 11q, indicating the presence of independent 11q-aUPD clones arising from a common clone having a heterozygous H396R mutation. (**c**). Taking the UPD proportions into consideration, chromatograms showing *c–Cbl* mutations are compatible with homozygous mutations except for that in case MDS.U189, MDS.U342, and MDS.U337, in which *c-Cbl* mutations were heterozygous without 11qUPD. **d**, CNAG outputs in two samples showing 11q-aUPD but with no detectable *c-Cbl* mutations. **e**, Somatic origin of mutations was confirmed in three cases, where mutated sequences in bone marrow specimens are compared with those in normal genomic DNA.

## Supplementary Figure 7




**Effect of c-Cbl mutation status on overall survival in CMML cases**
Overall survival of the patients with or without c-Cbl mutations was estimated by Kaplan-Meier method. The hazard ratio of c-Cbl mutated/non-mutated with the 95% confidential interval (95%CI) was calculated according to the proportional hazard model. The status of c-Cbl mutations does not seem to significantly affect overall suvivals (p = 0.75, Logrank test).

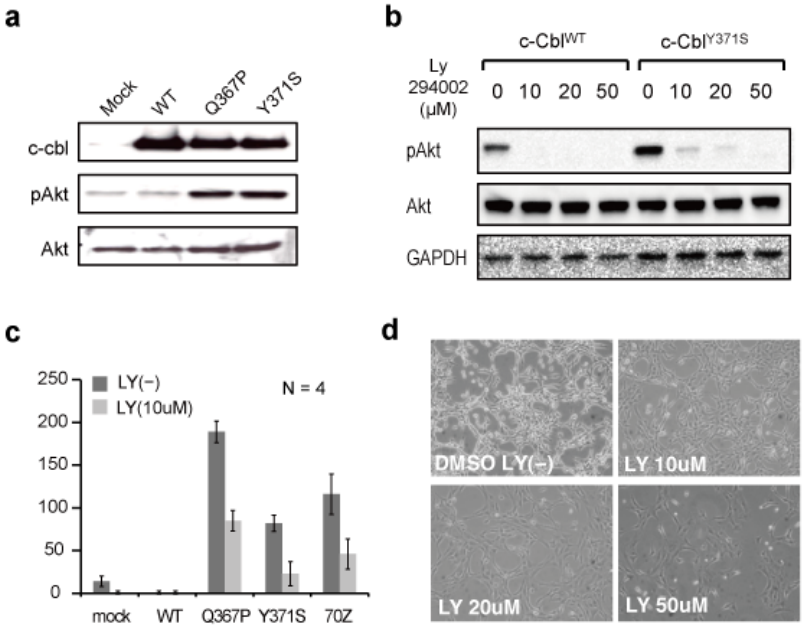
## Supplementary Figure 8




**Oncogenic potentials of mutant c-Cbl**

**a**, Representative pictures of cell morphology of NIH3T3 cells stably expressing c-Cbl mutants and their colony formation in soft-agar. Cell morphology (×100) and the results of colony assays are shown for NIH3T3 clones stably transduced with c-Cbl mutants (Q367P and C348G) together with wild-type c-Cbl and c-Cbl 70Z mutant. **b**, Stable expression of c-Cbl mutants in NIH3T3 cells were confirmed by Western blot analysis using anti-HA antibody. **c**, For colony assays, $1.0 \times 10^3$ NIH3T3 cells stably transduced with mock, wild–type c–Cbl, and 8 c–Cbl mutants isolated from myeloid neoplasms, as well as a human version of c–Cbl 70Z, were seeded in 0.33% soft agar supplemented with 10% FCS. The mean numbers (± s.d., N = 8) of colonies more than 1.0 mm in diameter are plotted. Significance of difference in colony counts as compared with mock-transduced cells were tested for each mutant using Mann-Whitney U test and p-values are provided. **d**, Tumor formation in nude mice, where $1.0 \times 10^7$ NIH3T3 cells expressing each *c–Cbl* mutant were injected subcutaneously in 3 mice bilaterally. Tumor formation was evaluated 21 days after the inoculation. The numbers of tumor formation sites are shown below each panel. **e**, The mean size (maximum diameter) of tumors (±s.d., N = 6) is plotted for each c-Cbl mutant. Note that tumors frequently disseminated also into the peritoneal cavity with massive ascites.

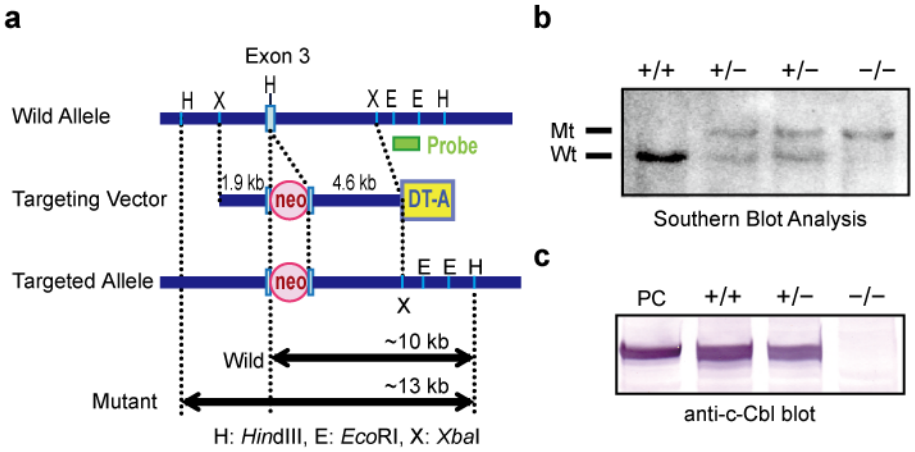
## Supplementary Figure 9




**Activation of Akt in NIH3T3**

**a**, Immunoblot analysis of total and phosphorylated Akt as well as c–Cbl showing actication of Akt pathway in NIH3T3 cells stably transduced with two *c–Cbl* mutants, *c-Cbl*$^{Q367P}$ and *c-Cbl*$^{Y371S}$, but not with mock and *c–Cbl*$^{WT}$. **b**, Immunoblot analysis of phosphorylated (top) and total (middle) Akt in NIH3T3 cells stably transfected with *c-Cbl*$^{WT}$ and *c-Cbl*$^{Y371S}$ treated with increasing doses of PI3K inhibitor (Ly294002). The bottom panel shows Anti-GAPDH blot. The Akt activation in *c-Cbl*$^{Y371S}$-transduced cells was inhibited by addition of Ly294002. **c**, The Effect of Ly294002 on anchorage-independent growth was evaluated by the colony formation assay in soft agar. Mean colony mubers (± s.d., N = 4) with or without Ly294002 are plotted for each c-Cbl mutant. Colony formation of mutant *c-Cbl*-transfected NIH3T3 cells was suppressed by addition of Ly294002. **d**, Transformed cell morphology of *c-Cbl*$^{Y371S}$-transduced NIH3T3 cells was reverted by treatment with increasing concentrations of Ly294002.

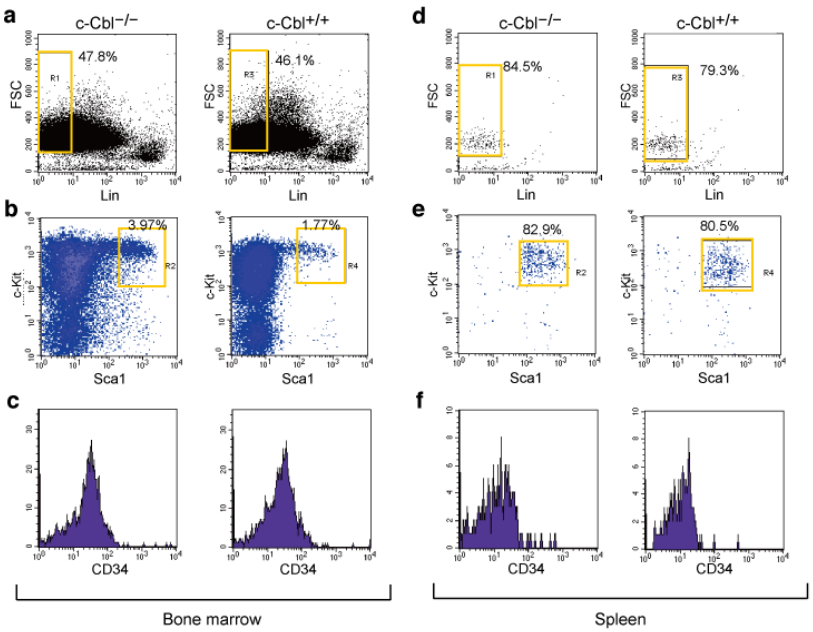
**Supplementary Figure 10**




**Generation of *c–Cbl* null mice by gene targeting**

To construct a targeting vector for homologous recombination, the MC1–neo sequence was inserted into the genomic fragment of the mouse c–Cbl gene to intervene c–Cbl exon 3 at the corresponding amino acid codon 191 and diphtheria toxin A (DT–A) sequence was added to the 3' terminal of the fragment in order to exclude non-homologous recombination events (**a**). The plasmid was linearized and electroporated into ES cell line CCE. The G418–resistant clones were screened for homologous recombination by Southern blot analysis. Two independent c–Cbl$^{+/-}$ ES clones were obtained, which were injected into blastocysts of C57BL/6 mice, and chimeras were obtained. Confirmation of germline transmission of the targeted allele in litter mates. Three c–Cbl genotypes, c–Cbl$^{+/+}$ (+/+), c–Cbl $^{+/-}$ (+/−) and c–Cbl$^{-/-}$ (−/−) animals were discriminated by 2 HindIII fragments representing a wild–type (Wt) allele (~10kb) and a mutant (Mt) allele (~13kb), as indicated (**b**). Loss of c–Cbl expression in c–Cbl$^{-/-}$ mouse were confirmed by Western blot analysis of mouse embryonic fibroblasts using the anti–mouse c–Cbl antibody. Total cell lysate from NIH3T3 transfected with c–Cbl was used as positive control (PC) (**c**).

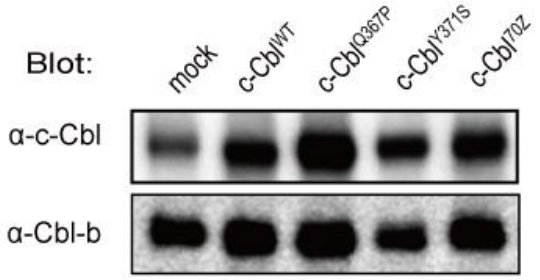
## Supplementary Figure 11




**Flow cytometric analysis of bone marrow cells and spleen cells from *c–Cbl*$^{+/+}$ and *c–Cbl*$^{-/-}$ mice**

Bone marrow and spleen cells were harvested from c-Cbl$^{+/+}$ and c-Cbl$^{-/-}$ mice and Lin− cells were purified using the MACS™ system and subjected to flow cytometry (**a**,**d**). After the MACS-purification, Lin− fractions fractions (R1and R3) were projected to identify and sort the Sca1+c–Kit+ fraction (LSK cells) (R2 and R4 in **b** and **e**). Each LKS population was further separated according to the CD34 expression (**c** and **f**). The proportion of R1-R4 fractions is provided in each panel.

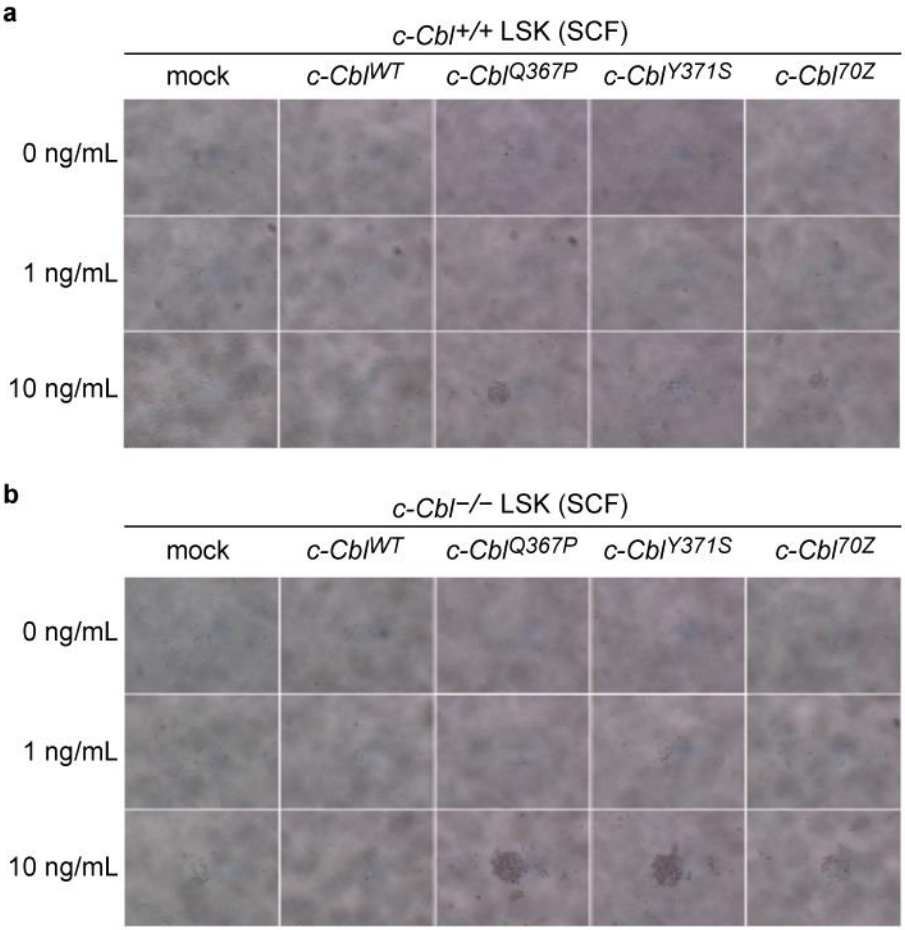
## Supplementary Figure 12




**Immunoblot analysis of expression of c-Cbl and Cbl-b in LSK HSPCs**

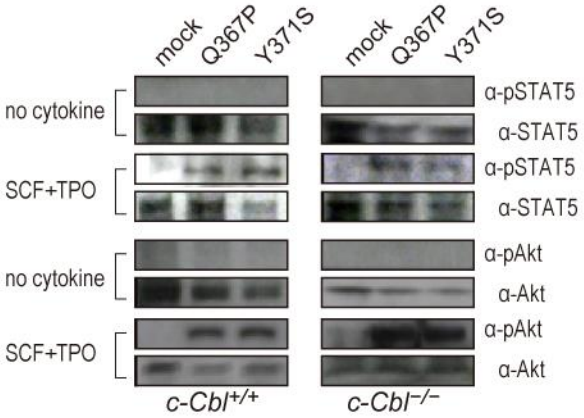
LSK cells from *c-Cbl*$^{+/+}$ mice were transduced with indicated *c-Cbl* constructs with an *IRES/EGFP* sequence. 5 x $10^4$ GFP-positive cells were analyzed for expression of c-Cbl and Cbl-b by immunoblot analysis with anti-c-Cbl and anti-Cbl-b antibodies, respectively.

**Supplementary Figure 13**




**Assay for sensitivity to SCF in LSK HSPCs transduced with c-Cbl mutants**
BM LSK HSPCs from *c-Cbl*$^{+/+}$ (**a**) and *c-Cbl*$^{-/-}$ (**b**) mice were infected with indicated retrovirus having IRES/EGFP. 50 GFP-positive cells were sorted into each of 96-well titer plates containing serum-free medium supplemented with SCF at different concentrations (0, 1, 10 ng/m). Photo-images of 5 replicate wells were taken after 5 days culture, and cell numbers were calculated by analyzing these photo-images.

**Supplementary Figure 14**




**Immunoblot analysis of activation of STAT5 and Akt in LSK HSPCs stimulated with SCF + TPO**
After 6h serum starvation, 1 × $10^4$ of *c-Cbl*$^{+/+}$ and *c-Cbl*$^{-/-}$ LSK cells transduced with various c-Cbl constructs were cultured in serum-free medium containing 10 ng/mL SCF and 10 ng/mL TPO for 15 min and whole cell lysates were analyzed by immunoblots using antibodies to STAT5, phosphorylated STAT5 (pSTAT5), Akt, and phosphorylated Akt (pAkt). Relative intensity of pSTAT5 and pAKt to total STAT5 and total Akt were calculated based on densitometry of the specific signals and plotted in Figure 4.

## Supplementary Reference


[31] Blake, T.J., Shapiro, M., Morse, H.C., 3rd *et al.*, The sequences of the human and mouse c-cbl proto-oncogenes show v-cbl was generated by a large truncation encompassing a proline-rich domain and a leucine zipper-like motif *Oncogene* 6, 653-657 (1991).

[32] Hamanaka, S., Nabekura, T., Otsu, M. *et al.*, Stable transgene expression in mice generated from retrovirally transduced embryonic stem cells *Mol Ther* 15, 560-565 (2007).

[33] Nabekura, T., Otsu, M., Nagasawa, T. *et al.*, Potent vaccine therapy with dendritic cells genetically modified by the gene-silencing-resistant retroviral vector GCDNsap *Mol Ther* 13, 301-309 (2006).

[34] Sanuki, S., Hamanaka, S., Kaneko, S. *et al.*, A new red fluorescent protein that allows efficient marking of murine hematopoietic stem cells *J Gene Med* 10, 965-971 (2008).

[35] Ory, D.S., Neugeboren, B.A., & Mulligan, R.C., A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes *Proc Natl Acad Sci U S A* 93, 11400-11406 (1996).

[36] Suzuki, A., Obi, K., Urabe, T. *et al.*, Feasibility of ex vivo gene therapy for neurological disorders using the new retroviral vector GCDNsap packaged in the vesicular stomatitis virus G protein *J Neurochem* 82, 953-960 (2002).

[37] Ema, H., Morita, Y., Yamazaki, S. *et al.*, Adult mouse hematopoietic stem cells: purification and single-cell assays *Nat Protoc* 1, 2979-2987 (2006).

[38] Osawa, M., Hanada, K., Hamada, H. *et al.*, Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell *Science* 273, 242-245 (1996).

*Nature* 478(7367):64-9. 参考論文

# Frequent pathway mutations of splicing machinery in myelodysplasia

Kenichi Yoshida[1†], Masashi Sanada[1†], Yuichi Shiraishi[2†], Daniel Nowak[7†], Yasunobu Nagata[1†], Ryo Yamamoto[3], Yusuke Sato[1], Aiko Sato-Otsubo[1], Ayana Kon[1], Masao Nagasaki[4], George Chalkidis[5], Yutaka Suzuki[6], Masashi Shiosaka[1], Ryoichiro Kawahata[1], Tomoyuki Yamaguchi[8], Makoto Otsu[3], Naoshi Obara[9], Mamiko Sakata-Yanagimoto[9], Ken Ishiyama[10], Hiraku Mori[11], Florian Nolte[7], Wolf-Karsten Hofmann[7], Shuichi Miyawaki[10], Sumio Sugano[6], Claudia Haferlach[12], H.Phillip Koeffler[13, 14], Lee-Yung Shih[15], Torsten Haferlach[12], Shigeru Chiba[9], Hiromitsu Nakauchi[3,8], Satoru Miyano[2, 5], and Seishi Ogawa[1]

[1]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan
[2]Laboratory of DNA Information Analysis, Human Genome Center, [3]Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, [4]Laboratory of Functional Genomics, Human Genome Center, [5]Laboratory of Sequence Data Analysis, Human Genome Center, and [6]Division of Systems Biomedical Technology, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[7]Department of Hematology and Oncology, University Hospital Mannheim, 1-3 Theodor-Kutzer-Ufer, Mannheim 68167, Germany
[8]Nakauchi Stem Cell and Organ Regeneration Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.
[9]Department of Clinical and Experimental Hematology, Institute of Clinical Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki, 305-8571, Japan
[10]Division of Hematology, Tokyo Metropolitan Ohtsuka Hospital, 2-8-1 Minami-Ohtsuka, Toshima-ku, Tokyo 170-0005, Japan
[11]Division of Hematology, Internal Medicine, Showa University Fujigaoka Hospital, 1-30 Fujigaoka, Aoba-ku, Yokohama, Kanagawa 227-8501, Japan
[12]Munich Leukemia Laboratory, Max-Lebsche-Platz 31, Munich, Germany
[13]Hematology/Oncology, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, CA 90048, USA
[14]National University of Singapore, Cancer Science Institute of Singapore, 28 Medical Drive, Singapore 117456
[15]Division of Hematology-Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University, 199 Tung Hwa North Rd, Taipei 105, Taiwan

Correspondence to:
Seishi Ogawa, M.D., Ph.D. (sogawa-tky@umin.ac.jp)

This paper was published in *Nature* 2011 Sep 11; 478(7367):64-9.
doi: 10.1038/nature10496.
http://www.nature.com/nature/journal/v478/n7367/full/nature10496.html

**Summary**

**Myelodysplastic syndromes and related myeloid neoplasms (myelodysplasia) are a heterogeneous group of myeloid neoplasms showing deregulated blood cell production with evidence of myeloid dysplasia, whose pathogenesis is still incompletely understood. Here we report a whole-exome sequencing study of 29 myelodysplasia specimens with matched germline controls. The spectrum of somatic mutations not only recapitulated our previous knowledge about the genetic pathways compromised in these neoplasms, but also revealed arrays of novel mutations, unexpectedly leading to the identification of frequent non-overlapping alterations in multiple components of the machineries involved in RNA splicing, including *U2AF35*, *ZRSR2*, *SRSF2*, and *SF3B1*. In a large series (N = 582) analysis, these spliceosome pathway mutations were frequent (~40 to ~90 % of the cases) in, and highly specific to myeloid neoplasms showing features of myelodysplasia. Conspicuously, most of the mutations affected genes involved in the 3'-splice site recognition during pre-mRNA processing, and thought to cause abnormal RNA splicing and compromised hematopoiesis. Our results provide the first evidence suggesting that genetic alterations of the splicing apparatus could be involved in human pathogenesis.**

Myelodysplastic syndromes (MDS) and related disorders (myelodysplasia) comprise a group of myeloid neoplasms characterized by deregulated, dysplastic blood cell production and a predisposition to acute myeloid leukemia (AML)[1,2]. While the prevalence of MDS has not been determined precisely, more than 10,000 people are estimated to develop myelodysplasia annually in the United States[3]. Their indolent clinical course before leukemic transformation and ineffective hematopoiesis with evidence of myeloid dysplasia indicate a pathogenesis distinct from that involved in *de novo* AML. Currently, a number of gene mutations and cytogenetic changes have been implicated in the pathogenesis of MDS, including mutations of *RAS*[4], *TP53*[5], *RUNX*[6,7], and more recently *ASXL1*[8], *c-CBL*[9-11], *IDH1/2*[12,13], *TET2*[14,15] and *EZH2*[16,17]. Nevertheless, mutations of this set of genes do not fully explain the pathogenesis of MDS because they are also found in other myeloid malignancies and roughly 20% of MDS cases have no known genetic changes (Ref.10 and also unpublished data). In particular, the genetic alterations responsible for the dysplastic phenotypes and ineffective hematopoiesis of myelodysplasia are poorly understood. Meanwhile, the recent development of massively parallel sequencing technologies has provided an expanded opportunity to discover genetic changes across the entire genomes or protein-coding sequences in human cancers at a single nucleotide level[18-23], which could be successfully applied to the genetic analysis of myelodysplasia to obtain a better understanding of its pathogenesis.

**Spectrum of somatic mutations in myelodysplasia**

In this study, we performed whole-exome sequencing of paired tumor/control DNA from 29 patients with MDS or related myeloid neoplasms (Supplementary Table 1). Although

incapable of detecting non-coding mutations and gene rearrangements, the whole-exome approach is a well-established strategy for obtaining comprehensive registries of protein-coding mutations at low cost and high performance. With the mean coverage of 130, 80.9 % of the target sequences were analyzed at more than x20 depth on average (Supplementary Fig. 1). All the candidates for somatic mutations (N = 497) generated through our data analysis pipeline were subjected to validation using Sanger sequencing (Supplementary Methods I and Supplementary Fig. 2). Finally, 268 non-synonymous somatic mutations were confirmed with an overall true positive rate of 53.9 % (Supplementary Fig. 3), including 206 missense, 25 nonsense, and 10 splice site mutations, and 27 frameshift-causing insertions/deletions (indels) (Supplementary Fig. 4). The mutation rate of 9.2 (0 – 21) per sample was significantly lower than that in solid tumors (16.2-302)[22,24,25] and multiple myeloma (28.4)[23] but was comparable to that in acute myeloid leukemia (7.3)[13,26,27] and chronic lymphocytic leukemia (11.5)[28]. Combined with the genomic copy number profile obtained by single nucleotide polymorphism (SNP) array karyotyping, this array of somatic mutations provided a landscape of myelodysplasia genomes (Ref. 29, 30 and Supplementary Fig. 5).

**Novel gene targets in myelodysplasia**

The list of the somatic mutations (Supplementary Table 2) included most of the known gene targets in myelodysplasia at similar mutation frequencies to those previously reported, indicating an acceptable sensitivity in our study. The mutations of the known gene targets, however, accounted for only 12.3 % of all detected mutations (N = 33), and the remaining 235 mutations involved previously unreported genes. Among these, nonsense/frameshift mutations are likely to compromise protein functions in most cases,

and their targets are possible candidates for tumor suppressors (Supplementary Table 3). Also, the involvement of common gene pathways previously implicated in myelodysplasia would argue for their causative roles rather than being indicative of passenger events. Thus, the mutations of genes in the *TP53*-related pathway (*DAPK1* and *TP53BP1*) and genes implicated in chromatin regulation (*DOT1-L, PHF6* and *PHF8*) are plausible candidates for driver mutations. For example, *Dot1-L*-deficient mice developed severe anemia attributable to defective erythropoiesis in bone marrow[31], whereas loss-of-function *PHF6* mutations have been reported in ~30 % of T-cell acute lymphoblastic leukemia (ALL)[32] and in our extended case series, 11 *PHF6* mutations were identified in 10 of 164 cases with myelodysplasia (Supplementary Fig. 6). Genes mutated in multiple cases are candidate targets of particular interest, for which high mutation rates are expected in the general population of myelodysplasia. In fact, 8 of the 12 recurrently mutated genes were among the well-described gene targets in myelodysplasia. The recurrent inactivation mutations of the *BCOR* (Xp11.4) gene, the responsible gene for the Lenz microphthalmia and oculofaciocardiodental syndromes, have not been reported previously in myelodysplasia, although *BCOR/RARα* fusion gene has been described in a rare case of AML[33]. Although BCOR has been shown to interact with histone deacetylases and to be involved in the repression of BCL6, probably through chromatin regulation, its functional impact on myeloid leukemogenesis remains to be elucidated. Similarly, the recurrent mutations of the remaining 3 genes, *U2AF35* (also known as *U2AF1*), *ZRSR2*, and *SRSF2* (*SFRS2, SC35*), were previously unreported in MDS and other myeloid neoplasms. They belong to a common pathway known as the machinery for RNA splicing or spliceosome (Table 1). Conspicuously, including additional 3 genes mutated in single cases (*SF3A1*, *SF3B1* and *PRPF40B*), 6 components

of the splicing machinery were mutated in 16 out of the 29 cases (55.2%) in a mutually exclusive manner (Fig. 1a, Supplementary Fig. 7 and Supplementary Table 2).

**Frequent mutations in splicing machinery in myelodysplasia**

RNA splicing is accomplished by a well-ordered recruitment, rearrangement and/or disengagement of a set of small nuclear ribonucleoprotein (snRNP) complexes (U1, U2, and either U4/5/6 or U11/12) as well as many other protein components onto the pre-mRNAs. What immediately drew our attention was the fact that the mutated components of the spliceosome are all engaged in the initial steps of RNA splicing, except for PRPF40B, whose functions in RNA splicing are poorly defined. Making physical interactions with SF1 and a SR protein, such as SRSF1 or SRSF2, the U2 auxiliary factor (U2AF) that consists of the U2AF65 (U2AF2)/U2AF35 heterodimer, is involved in the recognition of the 3'SS and its nearby polypyrimidine tract, which is thought to be required for the subsequent recruitment of the U2 snRNP, containing SF3A1 as well as SF3B1, to establish the splicing A complex (Fig. 1a)[34]. ZRSR2, also known as Urp (U2AF35-related protein), is another essential component of the splicing machinery. Showing a close structural similarity to U2AF35, ZRSR2 physically interacts with U2AF65 as well as SRSF1 and SRSF2, playing a distinct role from its homologue, U2AF35[35].

To confirm and extend the initial findings in the whole-exome sequencing, we studied mutations of the above 6 genes together with 3 additional spliceosome-related genes, including *U2AF65*, *SF1* and *SRSF1*, in a large series of myeloid neoplasms (N = 582) using a high-throughput mutation screen of pooled DNA followed by confirmation/identification of candidate mutations (Ref. 36, 37 and Supplementary

Methods II).

In total, 219 mutations were identified in 209 out of the 582 specimens of myeloid neoplasms through validating 313 provisional positive events in the pooled DNA screen (Supplementary Tables 4 and 5). The mutations among 4 genes, *U2AF35* (N = 37), *SRSF2* (N = 56), *ZRSR2* (N = 23) and *SF3B1* (N = 79), explained most of the mutations with a much lower mutational rate for *SF3A1* (N = 8), *PRPF40B* (N = 7), *U2AF65* (N = 4) and *SF1* (N = 5) (Fig. 1b). Mutations of the splicing machinery were highly specific to diseases showing myelodysplastic features, including MDS either with (84.9%) or without (43.9%) increased ring sideroblasts, chronic myelomonocytic leukemia (CMML) (54.5%), and therapy-related AML or AML with myelodysplasia-related changes (25.8%), but were rare in *de novo* AML (6.6%) and myeloproliferative neoplasms (MPN) (9.4%) (Fig. 2a). The mutually exclusive pattern of the mutations in these splicing pathway genes was confirmed in this large case series, suggesting a common impact of these mutations on RNA splicing and the pathogenesis of myelodysplasia (Fig. 2b). Meanwhile, the frequencies of mutations showed significant differences across disease types. Surprisingly, *SF3B1* mutations were found in the majority of the cases with MDS characterized by increased ring sideroblasts, i.e. refractory anemia with ring sideroblasts (RARS) (19/23 or 82.6%) and refractory cytopenia with multilineage dysplasia with ≥ 15 % ring sideroblasts (RCMD-RS) (38/50 or 76 %) with much lower mutation frequencies in other myeloid neoplasms. RARS and RCMD-RS account for 4.3 % and 12.9 % of MDS cases, respectively, where deregulated iron metabolism has been implicated in the development of refractory anemia[38]. With such high mutation frequencies and specificity, the *SF3B1* mutations were thought to be almost pathognomonic to these MDS subtypes characterized by increased ring sideroblasts, and

strongly implicated in the pathogenesis of MDS in these categories. Less conspicuously but significantly, *SRSF2* mutations were more frequent in CMML cases (Fig. 2a and Supplementary Table 4). Thus, while commonly involving the E/A splicing complexes, different mutations may still have different impacts on cell functions, contributing to the determination of discrete disease phenotypes. For example, studies have demonstrated that *SRSF2* was also involved in the regulation of DNA stability and that depletion of SRSF2 can lead to DNA hypermutability[39]. Of interest in this context, regardless of disease subtypes, samples with *SRSF2* mutations were shown to have significantly more mutations of other genes compared with *U2AF35* mutations ($p$ = 0.001, multiple regression analysis) (Supplementary Methods and Supplementary Fig. 8).

Notably, with a rare exception of A26V in a single case, the mutations of U2AF35 exclusively involved 2 highly conserved amino acid positions (S34 or Q157) within the N- and the C-terminal zinc finger motifs flanking the UHM domain. SRSF2 mutations exclusively occurred at P95 within an intervening sequence between the RPM and RS domains (Supplementary Figs. 9 and 10). Similarly, SF3B1 mutations predominantly involved K700E and, to a lesser extent, K666, H662 and E622, which are also conserved across species (Supplementary Fig. 11). The involvement of recurrent amino acid positions in these spliceosome genes strongly indicated a gain-of-function nature of these mutations, which has been a well-documented scenario in other oncogenic mutations, including RAS mutations involving codons 12, 13 and 61, as well as V617F JAK2, V600E BRAF, and more recently Y641 EZH2 mutations[40]. On the other hand, the 23 mutations in *ZRSR2* (Xp22.1) were widely distributed along the entire coding region (Fig. 1b). Among these, 14 mutations were nonsense or frameshift changes, or involved splicing donor/acceptor sites that caused either a premature truncation or a large

structural change of the protein, leading to loss-of-function. Combined with their strong male preference for the mutation (14/14 cases), *ZRSR2* most likely acts as a tumor suppressor gene with an X-linked recessive mode of genetic action. The remaining 9 *ZRSR2* mutations were missense changes and found in both males (6 cases) and females (3 cases), whose somatic origin was only confirmed in 2 cases. However, neither the dbSNP database (build131 and 132) nor the 1000 Genomes database (May 2011 snp calls) contained these missense nucleotides, suggesting that many, if not all, of these missense SNVs are likely to represent functional somatic changes, especially those found in males. Interrogation of these hot spots for mutations in *U2AF35* and *SRSF2* found no mutations among lymphoid neoplasms, including ALL (N = 24) or non-Hodgkin's lymphoma (N = 87) (data not shown).

**Functional implication of spliceosome mutations in deregulated RNA splicing**

Since the spliceosome mutations in myelodysplasia widely and specifically affect the major components of the splicing complexes E/A in a mutually exclusive manner, the common consequence of these mutations is logically the impaired recognition of 3'SSs that would lead to the production of aberrantly spliced mRNA species. To appreciate this and also to gain an insight into the biological/biochemical impact of these splicing mutations, we expressed the wild-type and the mutant (S34F) U2AF35 in HeLa cells using retrovirus-mediated gene transfer with EGFP marking (Fig. 3a) and examined their effects on gene expression in these cells using GeneChip® Human genome U133 plus 2.0 arrays, followed by gene set enrichment analysis (GSEA)[41]. In the current GSEA, all expressed genes were first ordered by the size of difference in their expression between wild-type and mutant U2AF35-transduced cells, and then the enriched gene sets were

explored. Intriguingly, the GSEA disclosed a significant enrichment of the genes on the nonsense-mediated mRNA decay (NMD) pathway among the significantly upregulated genes in mutant *U2AF35*-transduced HeLa cells (Fig. 3b, Supplementary Fig. 12a and Supplementary Table 6), which was confirmed by quantitative polymerase chain reactions (Fig. 3c). A similar result was also observed for the gene expression profile of a MDS-derived cell line (TF-1) transduced with the S34F mutant (Supplementary Figs. 12b-c). It is of note that the NMD activation by the mutant U2AF35 was suppressed significantly by the co-overexpression of the wild-type protein (Supplementary Fig. 12d), suggesting that the effect of the mutant protein is likely to be mediated by inhibition of the functions of the wild-type protein. Given that the NMD pathway, known as mRNA surveillance, provides a post-transcriptional mechanism for recognizing and eliminating abnormal transcripts that prematurely terminate translation[42], the result of the GSEA analyses indicated that the mutant U2AF35 induced abnormal RNA splicing in HeLa and TF1 cells, leading to the generation of unspliced RNA species having a premature stop codon and induction of the NMD activity.

To confirm this, we next performed whole transcriptome analysis in these cells using the GeneChip®Human exon 1.0 ST Array (Affymetrix), in which we differentially tracked the behavior of 2 discrete sets of probes showing different level of evidence of being exons, i.e. “Core” (authentic exons) and “non-Core” (more likely introns) sets (Supplementary Methods III and Supplementary Fig. 13). As shown in Figure 3d, the Core and non-Core set probes were differentially enriched among probes showing significant difference in expression between wild-type and mutant-transduced cells (FDR = 0.01). The Core set probes were significantly enriched in those probes significantly down-regulated in mutant *U2AF35*-transduced cells compared with wild-type

*U2AF35*-transduced cells, whereas the non-Core set probes were enriched in those probes significantly up-regulated in mutant *U2AF35*-transduced cells (Fig. 3e). The significant differential enrichment was also demonstrated, even when all probe sets were included (Fig. 3f). Moreover, the significantly differentially expressed Core set probes tended to be up- and down-regulated in wild-type and mutant *U2AF35*-transduced cells compared with mock-transduced cells, respectively, and vice versa for the differentially expressed non-Core set probes (Fig. 3e). Combined, these exon array results indicated that the wild-type U2AF35 correctly promoted authentic RNA splicing, whereas the mutant U2AF35 inhibited this processes, rendering non-Core and therefore, more likely intronic sequences to remain unspliced.

To consolidate the results in the exon array analysis and also to show more direct evidence for the splicing abnormalities in mutant-transduced cells, we performed sequencing analysis of mRNAs extracted from HeLa cells in which expression of the wild-type and mutant (S34F) U2AF35 were induced by doxycycline. The direct counts of those reads corresponding to the Core and non-Core set probes recapitulated the differential enrichment of both probe sets between the two HeLa samples (Supplementary Fig. 14). The abnormal splicing in mutant U2AF35-transduced cells was more directly demonstrated by evaluating the read counts in different fractions, i.e. exons, introns, ad intergene regions and at exon/intron junctions. First, after adjusted by the total number of mapped reads, the wild-type U2AF35-transduced cells showed an increased read counts in the exon fraction, but reduced counts in other fractions, compared with mutant U2AF35-transduced cells (Fig. 3g). The reads from the mutant-transduced cells were mapped to broader genomic regions compared with those from the wild-type U2AF35-transduced cells, which were largely explained by non-exon reads (Fig. 3h).

Finally, the number of those reads that encompassed the authentic exon/intron junctions were significantly increased in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells (Fig. 3i and Supplementary Methods VI). These results clearly demonstrated that failure of splicing ubiquitously occurred in mutant U2AF35-transduced cells. A typical example showing the abnormal splicing in mutant-transduced cells and the list of significantly unspliced exons is demonstrated in Supplementary Fig. 15 and Supplementary Table 7.

**Biological consequence of U2AF35 mutations**

Finally, we examined the effects of compromised functions of the E/A splicing complexes by gene mutations on cell proliferation and apoptosis. First, TF-1 and HeLa cells were transduced with lentivirus constructs expressing either the S34F U2AF35 mutant or wild-type U2AF35 under a tetracycline-inducible promoter (Figs. 4a-b, and Supplementary Fig. 16a), and cell proliferation was examined after the induction of their expression. Unexpectedly, after the induction of gene expression with doxycycline, the mutant U2AF35-transduced cells, but not the wild-type U2AF35-transduced cells, showed reduced cell proliferation (Fig. 4c and Supplementary Fig. 16b) with a marked increase in the G2/M fraction (G2/M arrest) together with enhanced apoptosis as indicated by the increased sub-G1 fraction and annexin V positive cells (Figs. 4d-f). To confirm the growth suppressive effect of U2AF35 mutants *in vitro*, a highly purified hematopoietic stem cell population ($CD34^{-}c\text{-}Kit^{+}ScaI^{+}\ Lin^{-}$, $CD34^{-}KSL$) prepared from C57BL/6 (B6)-Ly5.1 mouse bone marrow[43] was retrovirally transduced with either the mutant (S34F, Q157P and Q157R) or wild-type *U2AF35,* or the mock constructs, each harboring the EGFP marker gene (Supplementary Fig. 17). The ability of these

transduced cells to reconstitute the hematopoietic system was tested in a competitive reconstitution assay. The transduced cells were mixed with whole bone marrow cells from B6-Ly5.1/5.2 F1 mice, transplanted into lethally irradiated B6-Ly5.2 recipients, and peripheral blood chimerism derived from GFP-positive cells was assessed 6 weeks after transplantation by flow cytometry. We confirmed that each recipient mouse received comparable numbers of GFP-positive cells among the different retrovirus groups by estimating the % EGFP positive cells and overall proliferation in transduced cells by *ex vivo* tracking. Also no significant difference was observed in their homing capacity to bone marrow as assessed by transwell migration assays (Supplementary Fig. 18). As shown in Figure 4g, the wild-type U2AF35-transduced cells showed a slightly higher reconstitution capacity than the mock-transduced cells. On the other hand, the recipients of the cells transduced with the various U2AF35 mutants exhibited significantly lower $GFP^+$ cell chimerism than those of either the mock- or the wild-type U2AF35-transduced cells, indicating a compromised reconstitution capacity of the hematopoietic stem/progenitor cells expressing the U2AF35 mutants. In summary, these mutants lead to loss-of-function of U2AF35 most likely by acting in a dominant negative fashion to the wild-type protein.

**Discussion**

Our whole-exome sequencing study unexpectedly unmasked a complexity of novel pathway mutations found in ~40 to ~90% of myelodysplasia patients depending on the disease subtypes, which affected multiple but distinctive components (splicing complex E/A) of the splicing machinery and, as such, demonstrated the unquestionable power of massively parallel sequencing technologies in cancer research[13,20,22,23].

The RNA splicing system comprises an essential cellular machinery, through which eukaryotes can achieve successful transcription and guarantee the functional diversity of their protein species using alternative splicing in the face of a limited number of genes[44]. Accordingly, the meticulous regulation of this machinery should be indispensable for the maintenance of cellular homeostasis[45]. The current discovery of frequent mutations of the splicing pathway in myelodysplasia, therefore, represents another remarkable example that illustrates how cancer develops by targetting critical cellular functions. It also provides an intriguing insight into the mechanism of 'cancer specific' alternative splicing, which have long been implicated in the development of cancer, including MDS and other hematopoietic neoplasms[46,47].

In myelodysplasia, the major targets of spliceosome mutations seemed to be largely confined to the components of the E/A splicing complex, among others to *SF3B1*, *SRSF2*, *U2AF35*, and *ZRSR 2*, and to a lesser extent, to *SF3A1*, *SF1*, *U2AF65*, and *PRPF40B*. The broad coverage of the wide spectrum of spliceosome components in our exome sequencing was likely to preclude frequent involvement of other components on this pathway (Supplementary Fig. 19). The surprising frequency and specificity of these mutations in this complex, together with the mutually exclusive manner they occurred, unequivocally indicate that the compromised function of the E/A complex is a hallmark of this unique category of myeloid neoplasms, playing a central role in the pathogenesis of myelodysplasia. The close relationship between the mutation types and unique disease subtypes also support their pivotal roles in MDS.

Given the critical functions of the E/A splicing complex on the precise 3'SS recognition, the logical consequence of these relevant mutations would be the impaired splicing involving diverse RNA species. In fact, when expressed in HeLa cells, the

mutant U2AF35 induced global abnormalities of RNA splicing, leading to increased production of transcripts having unspliced intronic sequences, which are also supported the activated NMD pathway in mutant expressing cells. On the other hand, the functional link between the abnormal splicing of RNA species and the phenotype of myelodysplasia is still unclear. Mutant U2AF35 seemed to suppress cell growth/proliferation and induce apoptosis rather than confer a growth advantage or promote clonal selection. ZRSR2 knockdown in HeLa cells has been reported to also result in reduced viability, arguing for the common consequence of these pathway mutations[48]. These observations suggested that the oncogenic actions of these splicing pathway mutations are distinct from what is expected for classical oncogenes, such as mutated kinases and signal transducers, but could be more related to cell differentiation. Of note in this regard, the commonest clinical presentation of MDS is severe cytopenia in multiple cell lineages due to ineffective hematopoiesis with increased apoptosis rather than unlimited cell proliferation[1]. In this regard, lessons may be learned from the recent findings on the pathogenesis of the 5q- syndrome, where haploinsufficiency of *RPS14* leads to increased apoptosis of erythroid progenitors, but not myeloproliferation [49,50].

A lot of issues remain to be answered, however, to establish the functional link between these splicing pathway mutations and the pathogenesis of MDS, where the broad spectrum of RNA species affected by impaired splicing hampers identification of responsible gene targets. Moreover, the mutated components of the splicing machinery have distinct function of their own other than direct regulation of RNA splicing, involved in elongation and DNA stability, which may be important to determine specific disease phenotypes. Clearly, more studies are required to answer these questions through understanding of the molecular basis of their oncogenic actions.

## METHODS SUMMARY

Whole-exome sequencing of paired tumor/normal DNA samples from the 29 patients was performed after informed consent was obtained, as described in Supplementary Methods. SNP array-based analysis was performed as previously described, where genome-wide copy number alterations were called using the hidden Markov model-base algorithm implemented in CNAG software[29,30]. Gene mutations of target genes in a set of 582 myeloid neoplasms were performed by first screening mutations in PCR-amplified pooled targets from 12 individuals, followed by validation / identification of the candidate mutations within the corresponding 12 individuals by Sanger sequencing. For functional assays, FLAG-tagged cDNAs of the wild-type and mutant *U2AF35* were generated by *in vitro* mutagenesis, constructed into a MSCV-based retroviral vector, pGCDNsamIRESEGFP and a tetracycline-inducible lentivirus-based expression vector, LV-TRE-PRE-Ubc-tTA-I2G, and used for retrovirus- or lentivirus-mediated gene transfer to cell lines as well as $CD34^{-}$KSL cells. For gene expression studies, wild-type and mutant U2AF35 were expressed in HeLa cells and TF-1 cells by retrovirus-mediated gene transfer, from which RNA was extracted and analyzed on GeneChip®Human genome U133 plus 2.0 and Human exon 1.0 ST Arrays (Affymetrix), according to the manufacturer's instructions. Up-regulation of the NMD pathway genes in mutant U2AF35-transduced cells as found in microarray experiments, was validated by quantitative reverse transcriptase PCR. RNA sequencing was performed according to the manufacturer's instructions and analyzed for aberrant splicing (Supplementary Methods). Effects of U2AF35 mutants on cell proliferation were examined by MTT assays in HeLa and TF-1 cells transduced with lentivirus U2AF35 constructs as described, in which expression of wild-type and mutant U2AF35 were induced by doxycycline. For

competitive reconstitution assays, $CD34^-$ KSL cells harvested from C57BL/6 (B6)-Ly5.1 mice were retrovirally transduced using either the EGFP-only mock construct, the wild-type- or one of the three mutant-*U2AF35* constructs, each harboring the EGFP marker gene, and transplanted with competitor cells (B6-Ly5.1/5.2 F1 mouse origin) into lethally irradiated B6-Ly5.2 mice 48 hours after transduction. Frequency of GFP-positive cells was assessed in peripheral blood leukocytes by flow cytometry 6 weeks after transplantation. The primer sets used to validate gene mutations and qPCR assays of NMD gene expression are listed in Supplementary Tables 8−10, respectively. This study was approved by the ethics boards of the University of Tokyo, Munich Leukemia Laboratory, University Hospital Mannheim, University of Tsukuba, Tokyo Metropolitan Ohtsuka Hospital and Chang Gung Memorial Hospital.

**References**


1 Corey, S. J. *et al.* Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nat Rev Cancer* **7**, 118-129, doi:nrc2047 [pii]
10.1038/nrc2047 (2007).

2 Bejar, R., Levine, R. & Ebert, B. L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J Clin Oncol* **29**, 504-515, doi:JCO.2010.31.1175 [pii]
10.1200/JCO.2010.31.1175 (2011).

3 Ma, X., Does, M., Raza, A. & Mayne, S. T. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer* **109**, 1536-1542, doi:10.1002/cncr.22570 (2007).

4 Hirai, H. *et al.* A point mutation at codon 13 of the N-ras oncogene in myelodysplastic syndrome. *Nature* **327**, 430-432, doi:10.1038/327430a0 (1987).

5 Sugimoto, K. *et al.* Mutations of the p53 gene in myelodysplastic syndrome (MDS) and MDS-derived leukemia. *Blood* **81**, 3022-3026 (1993).

6 Song, W. J. *et al.* Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* **23**, 166-175, doi:10.1038/13793 (1999).

7 Osato, M. Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia. *Oncogene* **23**, 4284-4296, doi:10.1038/sj.onc.1207779
1207779 [pii] (2004).

8 Gelsi-Boyer, V. *et al.* Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol* **145**, 788-800, doi:BJH7697 [pii]
10.1111/j.1365-2141.2009.07697.x (2009).

9 Grand, F. H. *et al.* Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. *Blood* **113**, 6182-6192, doi:blood-2008-12-194548 [pii]
10.1182/blood-2008-12-194548 (2009).

10 Sanada, M. *et al.* Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature* **460**, 904-908, doi:nature08240 [pii]
10.1038/nature08240 (2009).

11 Dunbar, A. J. *et al.* 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies. *Cancer Res* **68**, 10349-10357, doi:68/24/10349 [pii]
10.1158/0008-5472.CAN-08-2754 (2008).

12 Yoshida, K. *et al.* A nonsense mutation of IDH1 in myelodysplastic syndromes and related disorders. *Leukemia* **25**, 184-186, doi:leu2010241 [pii]

10.1038/leu.2010.241 (2011).
13 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066, doi:NEJMoa0903840 [pii]
10.1056/NEJMoa0903840 (2009).
14 Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-2301, doi:360/22/2289 [pii]
10.1056/NEJMoa0810069 (2009).
15 Langemeijer, S. M. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* **41**, 838-842, doi:ng.391 [pii]
10.1038/ng.391 (2009).
16 Nikoloski, G. *et al.* Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* **42**, 665-667, doi:ng.620 [pii]
10.1038/ng.620 (2010).
17 Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* **42**, 722-726, doi:ng.621 [pii]
10.1038/ng.621 (2010).
18 Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145, doi:nbt1486 [pii]
10.1038/nbt1486 (2008).
19 Metzker, M. L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46, doi:nrg2626 [pii]
10.1038/nrg2626 (2010).
20 Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72, doi:nature07485 [pii]
10.1038/nature07485 (2008).
21 Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40**, 722-729, doi:ng.128 [pii]
10.1038/ng.128 (2008).
22 Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473-477, doi:nature09004 [pii]
10.1038/nature09004 (2010).
23 Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467-472, doi:nature09837 [pii]
10.1038/nature09837 (2011).
24 Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469**, 539-542, doi:nature09639 [pii]
10.1038/nature09639 (2011).
25 Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809-813,

doi:nature08489 [pii]
10.1038/nature08489 (2009).
26 Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-2433, doi:10.1056/NEJMoa1005143 (2010).
27 Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* **43**, 309-315, doi:ng.788 [pii]
10.1038/ng.788 (2011).
28 Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101-105, doi:nature10113 [pii]
10.1038/nature10113 (2011).
29 Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* **65**, 6071-6079, doi:65/14/6071 [pii]
10.1158/0008-5472.CAN-05-0465 (2005).
30 Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am J Hum Genet* **81**, 114-126, doi:S0002-9297(07)62821-3 [pii]
10.1086/518809 (2007).
31 Feng, Y. *et al.* Early mammalian erythropoiesis requires the Dot1L methyltransferase. *Blood* **116**, 4483-4491, doi:blood-2010-03-276501 [pii]
10.1182/blood-2010-03-276501 (2010).
32 Van Vlierberghe, P. *et al.* PHF6 mutations in T-cell acute lymphoblastic leukemia. *Nat Genet* **42**, 338-342, doi:ng.542 [pii]
10.1038/ng.542 (2010).
33 Yamamoto, Y. *et al.* BCOR as a novel fusion partner of retinoic acid receptor alpha in a t(X;17)(p11;q12) variant of acute promyelocytic leukemia. *Blood* **116**, 4274-4283, doi:blood-2010-01-264432 [pii]
10.1182/blood-2010-01-264432 (2010).
34 Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718, doi:S0092-8674(09)00146-9 [pii]
10.1016/j.cell.2009.02.009 (2009).
35 Tronchere, H., Wang, J. & Fu, X. D. A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* **388**, 397-400, doi:10.1038/41137 (1997).
36 Calvo, S. E. *et al.* High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet* **42**, 851-858, doi:ng.659 [pii]
10.1038/ng.659 (2010).
37 Bevilacqua, L. *et al.* A population-specific HTR2B stop codon predisposes to severe impulsivity. *Nature* **468**, 1061-1066,

doi:nature09629 [pii]
10.1038/nature09629 (2010).
38 Haase, D. *et al.* New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* **110**, 4385-4395, doi:blood-2007-03-082404 [pii]
10.1182/blood-2007-03-082404 (2007).
39 Xiao, R. *et al.* Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol Cell Biol* **27**, 5393-5402, doi:MCB.00288-07 [pii]
10.1128/MCB.00288-07 (2007).
40 Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet* **42**, 181-185, doi:ng.518 [pii]
10.1038/ng.518 (2010).
41 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:0506580102 [pii]
10.1073/pnas.0506580102 (2005).
42 Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* **5**, 89-99, doi:10.1038/nrm1310 (2004).
43 Ema, H. *et al.* Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nat Protoc* **1**, 2979-2987, doi:nprot.2006.447 [pii]
10.1038/nprot.2006.447 (2006).
44 Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**, 741-754, doi:nrm2777 [pii]
10.1038/nrm2777 (2009).
45 Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* **21**, 708-718, doi:21/6/708 [pii]
10.1101/gad.1525507 (2007).
46 Pajares, M. J. *et al.* Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol* **8**, 349-357, doi:S1470-2045(07)70104-3 [pii]
10.1016/S1470-2045(07)70104-3 (2007).
47 David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* **24**, 2343-2364, doi:24/21/2343 [pii]
10.1101/gad.1973010 (2010).
48 Shen, H., Zheng, X., Luecke, S. & Green, M. R. The U2AF35-related protein Urp contacts the 3' splice site to promote U12-type intron splicing and the second step of U2-type intron splicing. *Genes Dev* **24**,

2389-2394, doi:24/21/2389 [pii]
10.1101/gad.1974810 (2010).
49 Ebert, B. L. *et al.* Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335-339, doi:nature06494 [pii]
10.1038/nature06494 (2008).
50 Barlow, J. L. *et al.* A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nat Med* **16**, 59-66, doi:nm.2063 [pii]
10.1038/nm.2063 (2010).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements**

This work was supported by Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology. pGCDNsamIRESEGFP vector was a gift from Dr. Masafumi Onodera. We thank Yuka Mori, Otoe Hagiwara, Maki Nakamura and N. Mizota for their technical assistance. We are also grateful to Kiyomi Ikeuchi and Miki Ueda for their continuous encouragement throughout the study.

**Author Contributions**

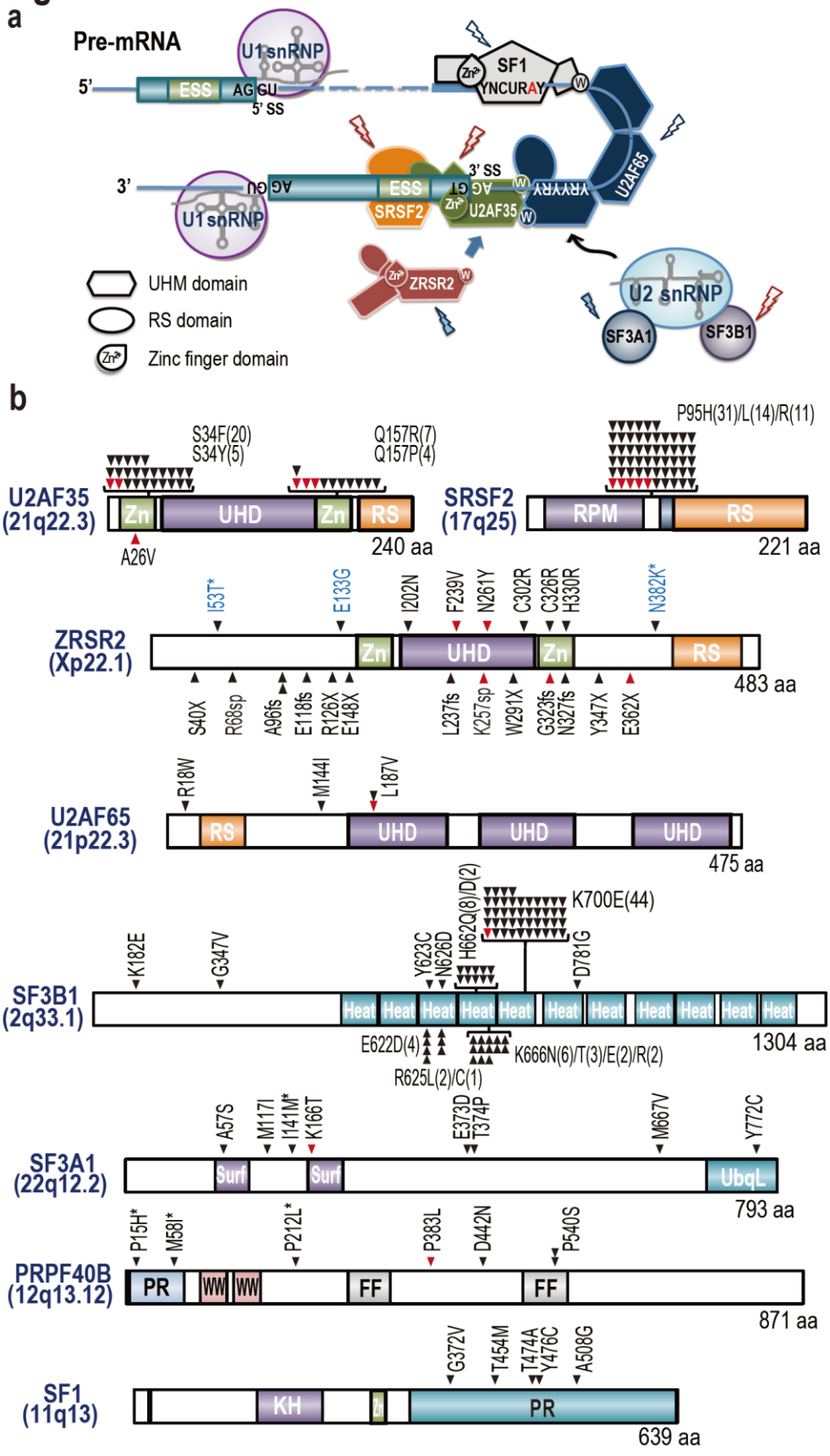
Y. Shiraishi, Y. Sato, A.S-O., Y.N., M.N., G.C., R.K., and S. Miyano were committed to bioinformatics analyses of resequencing data. M. Sanada, A.S-O., and Y. Sato, performed microarray experiments and their analyses. R.Y., T.Y., M.O., M. Sanada, A.K., M. Shiosaka, and H.N. involved in the functional analyses of U2AF35 mutants. N.O., M.S-Y., K.I., H.M., W-K.H., F.N., D.N., T.H., C.H., S. Miyawaki, S.C., H.P.K. and L-Y.S. collected specimens and were also involved in planning the project. K.Y., Y.N., Y. Suzuki, A.S-O., and S.S. processed and analyzed genetic materials, library preparation and sequencing. K.Y., M.Sanada., Y. Shiraishi, A.S-O., Y. Sato, and S.O. generated figures and Tables. S.O. lead the entire project and wrote the manuscript. All authors

participated in the discussion and interpretation of the data and the results.

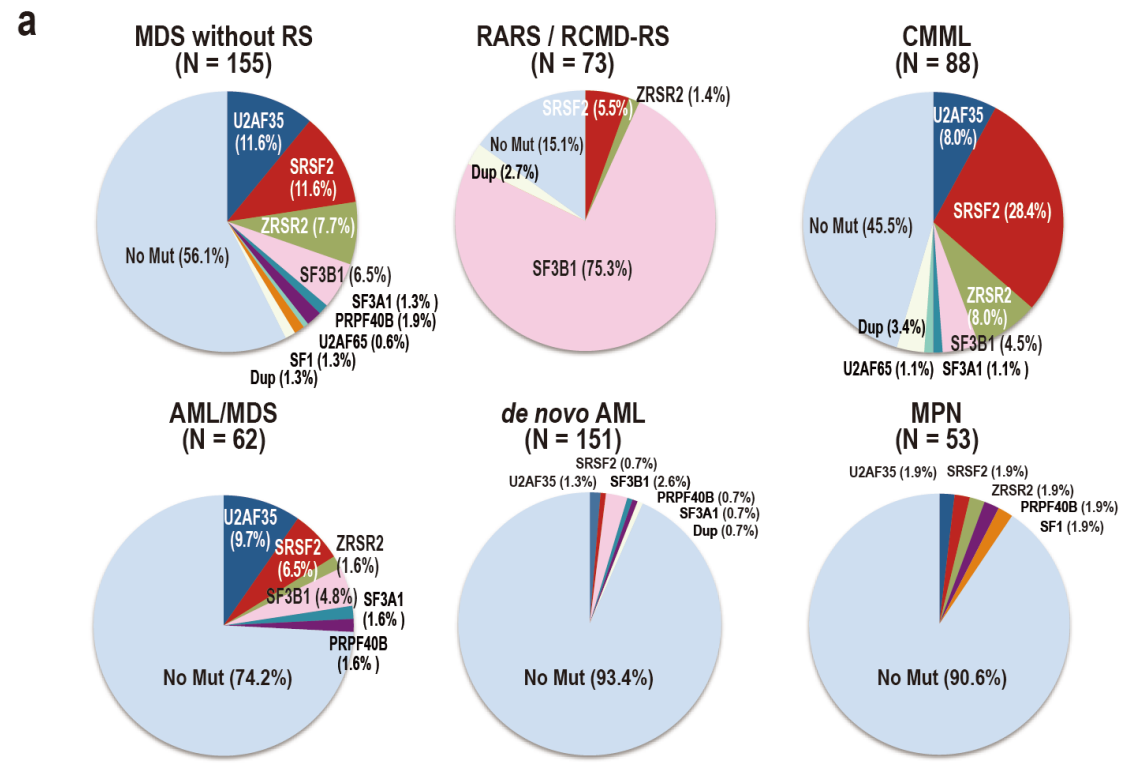
**Author Information**

Correspondence and requests for materials should be addressed to S.O. (sogawa-tky@umin.ac.jp).

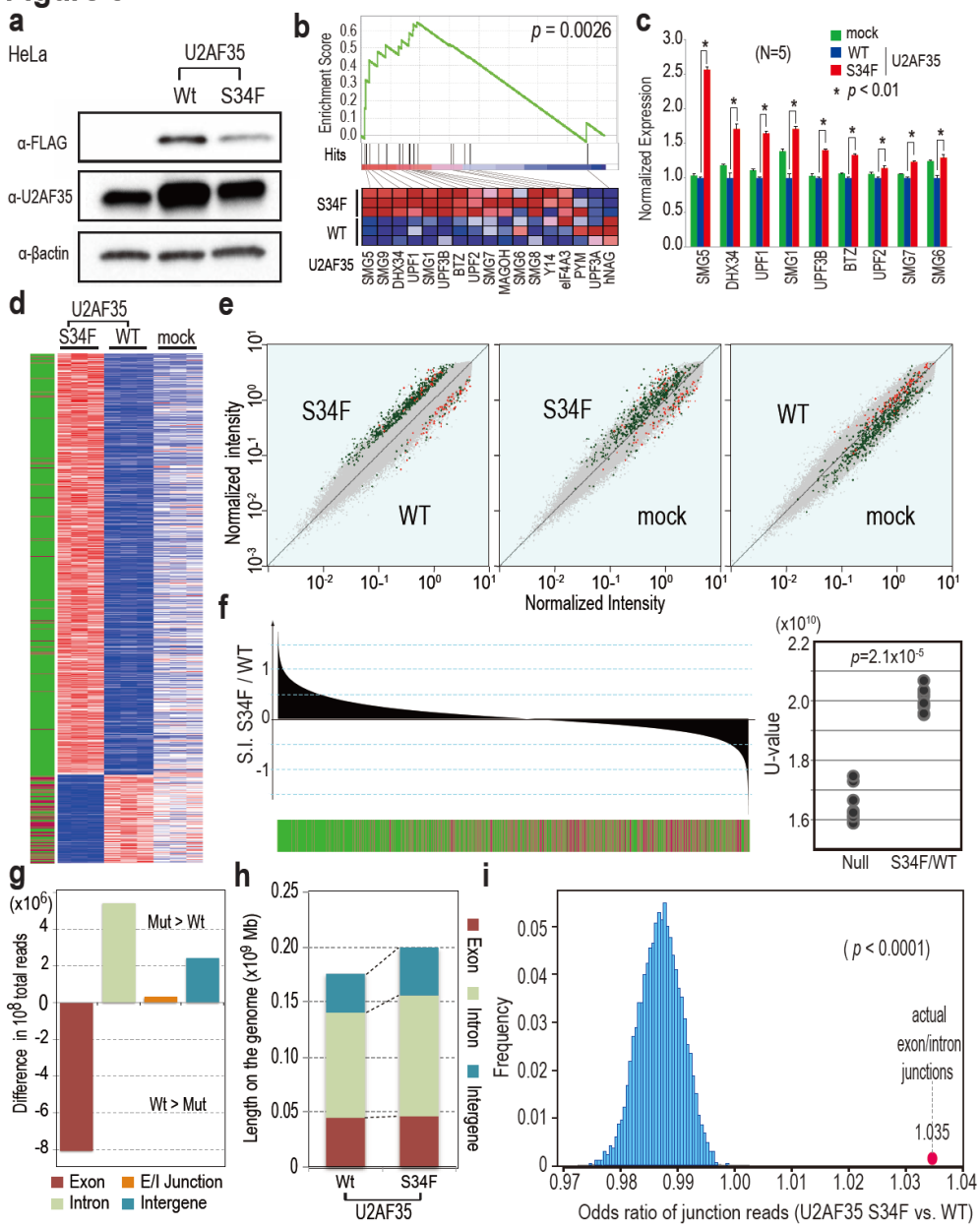
# Figure 1


a
Pre-mRNA
U1 snRNP
5'
ESS
AG GU
5' SS
SF1
YNCURAY
3'
3' SS
SRSF2
U2AF35
U2AF65
YRYYRY
ZRSR2
U2 snRNP
SF3A1
SF3B1
UHM domain
RS domain
Zinc finger domain
b
U2AF35 (21q22.3)
S34F(20)
S34Y(5)
Q157R(7)
Q157P(4)
Zn
UHD
RS
A26V
240 aa
SRSF2 (17q25)
P95H(31)/L(14)/R(11)
RPM
221 aa
ZRSR2 (Xp22.1)
I53T*
E133G
I202N
F239V
N261Y
C302R
C326R
H330R
N382K*
S40X
R68sp
A96fs
E118fs
R126X
E148X
L237fs
K257sp
W291X
G323fs
N327fs
Y347X
E362X
483 aa
U2AF65 (21p22.3)
R18W
M144I
L187V
475 aa
SF3B1 (2q33.1)
K182E
G347V
Y623C
N626D
H662Q(8)/D(2)
K700E(44)
D781G
Heat
E622D(4)
R625L(2)/C(1)
K666N(6)/T(3)/E(2)/R(2)
1304 aa
SF3A1 (22q12.2)
A57S
M117I
I141M*
K166T
E373D
T374P
M667V
Y772C
Surf
UbqL
793 aa
PRPF40B (12q13.12)
P15H*
M58I*
P212L*
P383L
D442N
P540S
PR
WW
FF
871 aa
SF1 (11q13)
G372V
T454M
T474A
Y476C
A508G
KH
PR
639 aa

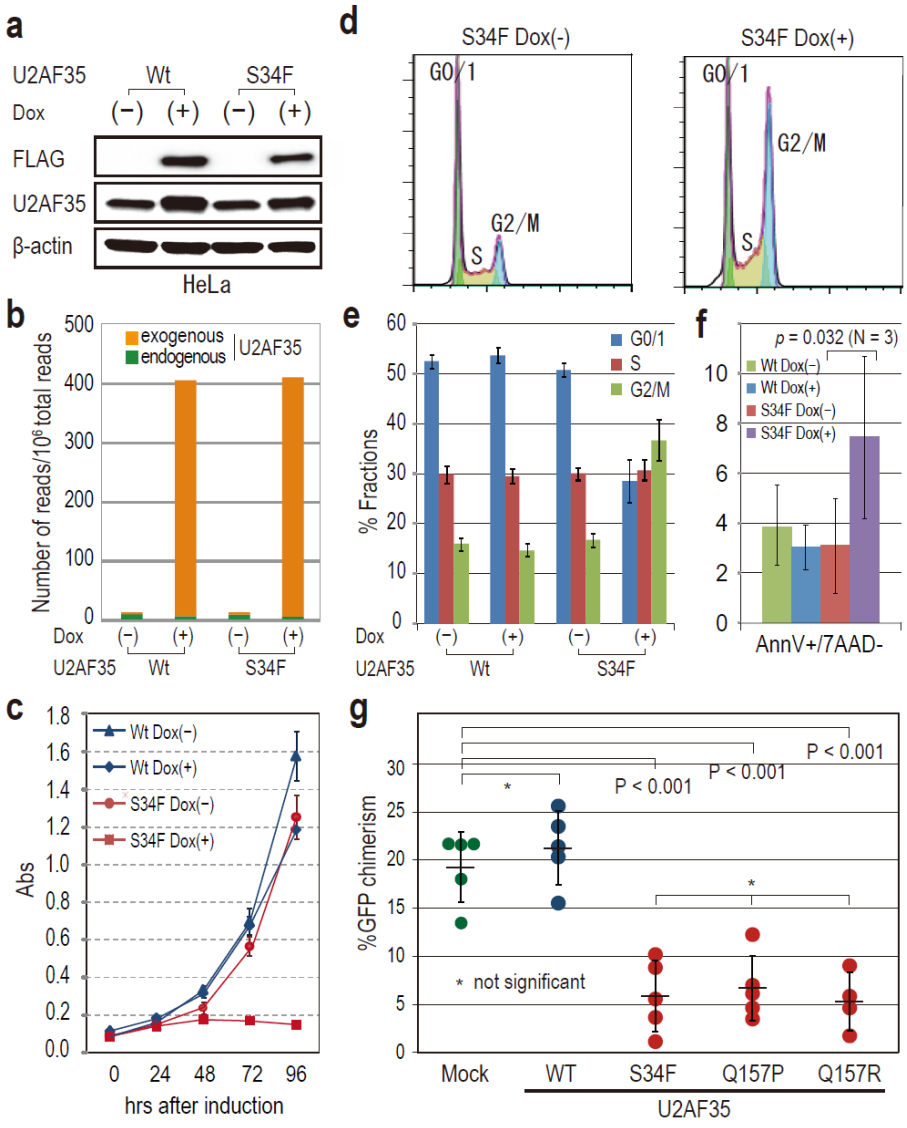
# Figure 2


a
MDS without RS
(N = 155)
U2AF35 (11.6%)
SRSF2 (11.6%)
ZRSR2 (7.7%)
No Mut (56.1%)
SF3B1 (6.5%)
SF3A1 (1.3% )
PRPF40B (1.9%)
U2AF65 (0.6%)
SF1 (1.3%)
Dup (1.3%)
RARS / RCMD-RS
(N = 73)
SRSF2 (5.5%)
ZRSR2 (1.4%)
No Mut (15.1%)
Dup (2.7%)
SF3B1 (75.3%)
CMML
(N = 88)
U2AF35 (8.0%)
SRSF2 (28.4%)
No Mut (45.5%)
ZRSR2 (8.0%)
Dup (3.4%)
SF3B1 (4.5%)
U2AF65 (1.1%)
SF3A1 (1.1% )
AML/MDS
(N = 62)
U2AF35 (9.7%)
SRSF2 (6.5%)
ZRSR2 (1.6%)
SF3B1 (4.8%)
SF3A1 (1.6% )
PRPF40B (1.6% )
No Mut (74.2%)
de novo AML
(N = 151)
SRSF2 (0.7%)
U2AF35 (1.3%)
SF3B1 (2.6%)
PRPF40B (0.7%)
SF3A1 (0.7%)
Dup (0.7%)
No Mut (93.4%)
MPN
(N = 53)
U2AF35 (1.9%)
SRSF2 (1.9%)
ZRSR2 (1.9%)
PRPF40B (1.9%)
SF1 (1.9%)
No Mut (90.6%)
b
RARS/RCMD-RS
MDS without RS
CMML
AML/MDS
de novo AML
MPN
SF3B1
SFRS2
U2AF35
ZRSR2
SF3A1
PRPF40B
U2AF65
SF1

# Figure 3


a
HeLa
U2AF35
Wt
S34F
α-FLAG
α-U2AF35
α-βactin
b
p = 0.0026
Enrichment Score
Hits
S34F
WT
U2AF35
SMG5
SMG9
DHX34
UPF1
SMG1
UPF3B
BTZ
UPF2
SMG7
MAGOH
SMG6
SMG8
Y14
eIF4A3
PYM
UPF3A
hNAG
c
(N=5)
mock
WT
S34F
U2AF35
* p < 0.01
Normalized Expression
SMG5
DHX34
UPF1
SMG1
UPF3B
BTZ
UPF2
SMG7
SMG6
d
U2AF35
S34F
WT
mock
e
S34F
WT
S34F
mock
WT
mock
Normalized intensity
Normalized Intensity
f
S.I. S34F / WT
(x10^10)
p=2.1x10^-5
U-value
Null
S34F/WT
g
(x10^6)
Mut > Wt
Wt > Mut
Difference in 10^8 total reads
Exon
E/I Junction
Intron
Intergene
h
Length on the genome (x10^9 Mb)
Wt
S34F
U2AF35
Exon
Intron
Intergene
i
( p < 0.0001)
actual exon/intron junctions
1.035
Frequency
Odds ratio of junction reads (U2AF35 S34F vs. WT)

# Figure 4


a
U2AF35
Wt
S34F
Dox
(−) (+) (−) (+)
FLAG
U2AF35
β-actin
HeLa
b
Number of reads/10^6 total reads
exogenous
endogenous
U2AF35
Dox
U2AF35
Wt
S34F
c
Abs
Wt Dox(−)
Wt Dox(+)
S34F Dox(−)
S34F Dox(+)
hrs after induction
d
S34F Dox(-)
S34F Dox(+)
G0/1
S
G2/M
e
% Fractions
G0/1
S
G2/M
f
p = 0.032 (N = 3)
Wt Dox(−)
Wt Dox(+)
S34F Dox(−)
S34F Dox(+)
AnnV+/7AAD-
g
%GFP chimerism
P < 0.001
P < 0.001
P < 0.001
* not significant
Mock
WT
S34F
Q157P
Q157R
U2AF35

## Figure Legends

### Figure 1. Splicing complex E/A and frequent gene mutations in their components

**a.** In the initial steps of RNA splicing, U1 snRNP is first recruited to the 5' SS of pre-mRNAs as they are transcribed, while SF1 and the larger subunit of the U2 auxiliary factor (U2AF), U2AF65, bind the branch point sequence (BPS) and its downstream polypyrimidine tract. The smaller subunit of U2AF (U2AF35) binds to the AG dinucleotide of the 3' SS, interacting with both U2AF65 and a SR protein, such as SRSF2 (SFRS2, SC35), through its UHD and SR domain, respectively, comprising the earliest splicing complex (E complex). ZRSR2, or Urp, also interacts with U2AF and SR proteins to play essential roles in RNA splicing. After the recognition of the 3' SS through these factors, U2 snRNP, a multiple component protein/snRNA complex including SF3A1 and SF3B1, is recruited to the 3' SS to generate the splicing complex A. In myelodysplasia, multiple components of these splicing complexes are affected by gene mutations, providing a novel and conspicuous example of 'pathway mutations' in human cancer. The mutated components identified through whole-exome sequencing are indicated by arrows.

**b.** Each mutation in the 8 spliceosome components of splicing complex E/A is shown with an arrowhead. Confirmed somatic mutations are discriminated by red arrows. Mutations which were predicted as SNPs by MutationTester (http://www.mutationtaster.org/) are indicated by asterisks. The number of each mutation is indicated in parenthesis. Mutations of *U2AF35* and *SRSF2* and *SF3B1*, involved distinct hotspots, while those of *ZRSR2* were widely distributed along the entire coding sequence, most of which were nonsense or splice site mutations, or indels, causing

premature truncation of the protein. *ZRSR2* mutations in females are shown in blue. Known domain structures are shown in colored boxes as indicated.

**Figure 2. Frequencies and distribution of spliceosome pathway gene mutations in myeloid neoplasms**

**a**. Frequencies of spliceosome pathway mutations among 582 cases with myeloid neoplasms including the initial 29 cases analyzed by whole-exome sequencing. Frequencies are shown for each type of neoplasms, i.e. MDS either with or without increased ring sideroblasts, CMML, AML/MDS including both AML with myelodysplasia-related changes and therapy-related AML, *de novo* AML and myeloproliferative neoplasms (MPN). **b**. Distribution of multiple spliceosome pathway mutations is shown for these 8 genes. Diagnosis of each sample is shown by indicated colors.

**Figure 3. Altered RNA-splicing caused by a *U2AF35* mutant**

**a.** Western blot analyses showing expression of transduced wild-type or mutant (S34F) U2AF35 in HeLa cells used for the analyses of expression and exon microarrays. **b**. The result of the GSEA demonstrating a significant enrichment of the set of 17 NMD pathway genes among significantly differentially expressed genes between wild-type and mutant U2AF35-transduced HeLa cells. The significance of the gene set was empirically determined by 1,000 gene-set permutations. **c**. The results of qPCR confirming the microarray results of the expression of 9 genes that contributed to the core enrichment in the NMD gene set. Means and S.E. from five experiments are shown for each NMD gene expression of mock, wild-type and mutant U2AF35-transduced HeLa cells, after

normalized to that of wild-type U2AF35-transduced HeLa cells. *P* values were determined by the Mann-Whitney *U* test. **d**. Significantly up-regulated and down-regulated probe sets (FDR = 0.01) in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells in triplicate exon array experiments are shown in a heat map. The origin of each probe set is depicted in the left lane, where red and green bars indicate the Core and non-Core sets, respectively. **e**. Pair-wise scatter plots of the normalized intensities of entire probe sets (grey) across different experiments. The Core and non-Core set probes that were significantly differentially expressed between the wild-type and mutant U2AF35-transduced cells are plotted in red and green, respectively. **f.** Distribution of the Core (red) and non-Core (green) probe sets within the entire probe sets ordered by splicing index (S.I., Supplementary Methods III), calculated between wild-type and mutant U2AF35-transduced cells. **g**. Difference in read counts for the indicated fractions per 1 x $10^8$ total reads in RNA sequencing between wild-type- and mutant U2AF35-expressing HeLa cells analysis. Increased/decreased read counts in mutant U2AF35-expressing cells are plotted upward/downward, respectively. **h**. Comparison of the genome coverage by the indicated fractions in wild-type and mutant U2AF35 expressing cells. The genome coverage was calculated for each fraction within the $10^8$ reads randomly selected from the total reads and averaged for 10 independent selections. **i.** The odds ratio of the junction reads within the total mapped reads was calculated between the two experiments (red circle), which was evaluated against the 10,000 simulated values under the null hypothesis (histogram in blue).

**Figure 4. Functional analysis of mutant U2AF35**

**a**. Western blot analyses of doxycycline-induced expression of mutant (S34F) and

wild-type U2AF35 in HeLa cells performed before and 72h after induction. Both anti-FLAG and anti-U2AF35 blots are shown to allow for the estimation of endogenous and exogenous protein levels. **b**. Evaluation of the relative contribution of endogenous and exogenous U2AF35 expression using massive sequencing of transcriptome, in which total number of reads containing the mutated position was differentially enumerated for the mutated nucleotide in each sample. **c**. Cell proliferation assays of U2AF35-transduced HeLa cells, where cell numbers were measured using Cell-counting kit and are plotted as their mean absorbance ±S.D. (N = 6). **d and e**. The flow cytometry analysis of PI-stained HeLa cells transduced with the S34F U2AF35 mutant, showing a prominent increase in the G2/M fraction after induction of the mutant with doxycycline. Indicated error bars mean S.D. (N=3). **f**. Fractions of the annexin V-positive population among the 7AAD-negative population plotted as their mean and S.D.(N=3) for the indicated samples before and after induction of wild-type and mutant U2AF35. The significant difference was determined by paired t-test. **g**. The result of competitive reconstitution assay for $CD34^{-}$KSL cells transduced with either of the 3 *U2AF35* mutants compared with mock- and wild-type transduced cells; the horizontal lines and vertical lines indicate the mean and S.D. of the results from 5 mice, respectively, where outliers were excluded from the analysis using the Grubbs test. The significance of differences was determined by Bonferroni's multiple comparison test. The vertical axis shows the % GFP-positive Ly5.1 cells in peripheral blood 6 weeks after transplantation.

**Supplementary Information**

# Frequent pathway mutations of splicing machinery in myelodysplasia

Kenichi Yoshida, Masashi Sanada, Yuichi Shiraishi, Daniel Nowak, Yasunobu Nagata, Ryo Yamamoto, Yusuke Sato, Aiko Sato-Otsubo, Ayana Kon, Masao Nagasaki, George Chalkidis, Yutaka Suzuki, Masashi Shiosaka, Ryoichiro Kawahata, Tomoyuki Yamaguchi, Makoto Otsu, Naoshi Obara, Mamiko Sakata-Yanagimoto, Ken Ishiyama, Hiraku Mori, Florian Nolte, Wolf-Karsten Hofmann, Shuichi Miyawaki, Sumio Sugano, Claudia Haferlach, H.Phillip Koeffler, Lee-Yung Shih, Torsten Haferlach, Shigeru Chiba, Hiromitsu Nakauchi, Satoru Miyano, and Seishi Ogawa

**Contents**

## Supplementary Methods

I. Whole-exome sequencing

II. Mutation analysis enhanced by deep sequencing of pooled target exons

III. Preparation of high-titer virus stock

IV. Whole transcriptome analysis using GeneChip®Human genome U133 plus 2.0 Arrays and Exon 1.0 ST Arrays

V. Quantitative RT PCR

VI. RNA sequencing

VII. Cell proliferation assays and assays for cell cycle and apoptosis analyses

VIII. Mouse KSL Competitive reconstitution assay

**Supplementary Figures**

Figure 1
Mean coverage of whole-exome sequencing in 29 paired tumor/control samples

Figure 2
The data processing pipeline for detecting somatic nonsynonymous mutations in whole-exome sequencing

Figure 3
True positive rates in Sanger-based validation of candidate somatic mutations

Figure 4
Number of somatic changes in 29 patients with myelodysplasia

Figure 5
Landscape of myelodysplasia genomes as revealed by whole-exome analysis combined with SNP array karyotyping

Figure 6
*PHF6* mutations in MDS and related myeloid neoplasms

Figure 7
Validation of somatic mutations in 6 spliceosome pathway genes detected by whole-exome sequencing by Sanger sequencing

Figure 8
Overlapping of spliceosome pathway gene mutations with common gene targets in myelodysplasia

Figure 9
Alignment of U2AF35 amino acid sequences of different species

Figure 10
Alignment of SRSF2 amino acid sequences of different species

Figure 11
Alignment of SF3B1 amino acid sequences of different species

Figure 12
Activation of the NMD pathway in Hela and TF-1 cells transduced with either wild-type and mutant U2AF35

Figure 13
The three discrete sets of probes on GeneChip®Human Exon 1.0 ST Array

Figure14
Differential enrichment of exon and intron reads reproducibly observed in RNA sequencing

Figure15
A typical example of abnormal splicing caused by mutant U2AF35 detected by exon arrays and RNA sequencing

Figure16
Suppression of cell growth of TF-1 cell line transduced with mutant U2AF35

Figure17
Expression of transduced U2AF35 in mouse $CD34^-$ KSL cells

Figure18
Transwell® migration assays of mouse progenitor cells transduced with U2AF35 mutants

Figure19
The mean depth of analysis for genes involved in RNA splicing

## Supplementary Tables

Table 1. Characteristics of 29 patients subjected to whole-exome analysis

Table 2. The list of the mutated genes in whole-exome analysis

Table 3. Nonsense and frameshift mutations detected in whole-exome analysis

Table 4. Spliceosome mutations in myeloid neoplasms

Table 5. Mutation analysis of splicesome mutations in 582 cases with myeloid neoplasms

Table 6. The gene sets for RNA export from nucleus and extended NMD genes

Table 7. List of significantly unspliced exonic regions

Table 8. List of primers used for validation of somatic mutations detected in whole-exome analysis

Table 9. List of PCR primers used for pooled target sequencing and validation

Table 10. Primer sets used for qPCR analysis of NMD pathway genes

**Supplementary Methods**

## I. Whole-exome sequencing

### 1. Sequencing procedures

Tumor DNAs were extracted from patients' bone marrow or peripheral blood mononuclear cells. For germline control, DNA was obtained from either buccal mucosa or paired CD3 positive T cells with or without prior culture in the presence of phytohemagglutinin and IL-2. Whole exome capture was accomplished based on liquid phase hybridization of sonicated genomic DNA having 150 - 200bp of mean length to the bait cRNA library synthesized on magnetic beads (SureSelect®, Agilent Technology), according to the manufacture's protocol. SureSelct Human All Exon 38Mb kit was used for the initial 16 cases, and 50Mb kit was available for the remaining 13 cases (Supplementary Table 1). The captured targets were subjected to massive sequencing using Illumina GAIIx and/or HiSeq 2000 with the pair end 75-108 bp read option, according to the manufacture's instruction.

### 2. Pipeline for data processing

The raw sequence data generated from Illumina GAIIx or HiSeq2000 sequencers were processed through the in-house pipeline constructed for whole-exome analysis of paired cancer genomes at Human Genome Center the Institute of Medical Science, University of Tokyo, which are summarized in Supplementary Figure 2. The data processing is divided into two steps,

1) Generation of a .bam file (http://samtools.sourceforge.net/) for paired normal and tumor samples for each case.

2) Detection of somatic point mutations and indels by comparing normal and tumor BAM files.

### 2.1 Generation of .bam files

#### 2.1.1 Preprocessing

First, .fastq files originally generated from Illumina sequencers are converted to .fastq in Sanger format via the maq-sol2sanger [http://maq.sourceforge.ned/]. PCR adaptor sequences contaminated in the sequence reads were removed by the following procedure: If the first 12 consecutive 3' bases of the opposite adaptor sequence were matched, the remaining bases were matched to the remaining adaptor sequence. If and only if the all the remaining sequence was completely matched, all the matched bases were removed from the read.

#### 2.1.2 Mapping of sequence reads and detection of duplicate reads

Sequenced reads are aligned to the NCBI Human Reference Genome Build 37 with BWA (version 0.5.8 and default parameter settings) [http://bio-bwa.sourceforge.net/]. The output is written into a .sam file, which were converted into a .bam file format for the subsequent calculations via SamFormatConverter in the Picard suit [http://picard.sourceforge.net/Picard]. The aligned reads are examined with the MarkDuplicates algorithm from Picard to identify molecular duplicates, where a read is considered a molecular duplicate, if both ends of the pair reads are mapped to the identical genomic locations. The detected duplicates are flagged in the .bam file.

2.1.3 Local re-alignment

Afrer mapping to the reference genome and detection of duplicate reads, local re-alignment was performed to increase the sensitivity and specificity of indel detection, in which the entire context of multiple mapped reads has to be taken into account and serves as evidence for putative indels. Most short read aligners map each read independently to the reference genome and hence reads supporting indels may be aligned with multiple mismatches to the reference rather than with a gap. We applied IndelRealigner in Genome Analysis Tool Kit (GATK) [http://www.broadinstitute.org/gatk] to perform multiple sequence re-alignment around the locations of candidate indels. This step tries to correct the above mentioned placement errors and decrease the false positive rate of indel and SNV calls.

2.1.4 Recalibration of base quality and aggregation of sample data in multiple lanes

Each sequenced nucleotide within a short read was associated with a Phred-like quality value which indicates the probability that the base call was wrong. We used GATK to recalibrate the base quality scores after the re-alignment process. The software recalculates Q scores based on the originally reported quality score, the position of the nucleotide within the read, and the preceding and current nucleotide. Finally, sequence data for the same sample from multiple lanes were combined into single .bam file using samtools-merge command [http://samtools.sourceforge.net/].

## 2.2 Detection of somatic mutations and indels

2.1.1 Generation of the pile-up files for tumor and control data

Before summarizing base call data, low quality reads are eliminated from each .bam file, including those reads which have more than 5 mismatches to the reference sequences or whose mapping quality was less than 30 were removed. The sequence data in .bam files were then summarized into .pileup file, which contains the counts of each base call at every nucleotide position in the target sequences. To suppress too many false positive finding, the following nucleotide positions were eliminated from the further analysis, including those positions at which the depth is less than 10 in either tumor or control, or the most frequent SNV or indel accounts less than 7% of all reads in tumor. The SNV comprising equal to or more than 7% of total reads at each nucleotide position, if exists, is adopted as the candidate mutation.

2.2.1 Statistical evaluation of SNVs and indels

The significance of each candidate mutation is evaluated by Fisher's exact test by enumerating the number of the reference base and the candidate SNV in both tumor and control. Candidate mutations either (i) having p-values of less than 0.001, or (ii) having p-values of less than 0.05, more than 20% of the mutated allele in tumor sample and less than 10% of the mutated all in normal sample, were adopted as provisional candidate for somatic mutations. The provisional candidates for somatic mutations were subjected to validation using Sanger sequencing, if they do not satisfy any of the following conditions:

- SNPs registered in dbSNP 131.
- SNPs found in the1000 Genomes [http://www.1000genomes.org/] database.
- Variants on the intron region excluding splicing sites.

・Synonymous mutations.

The primer set used to validate mutations are listed in Supplementary Table 8.

## II. Mutation analysis enhanced by deep sequencing of pooled target exons

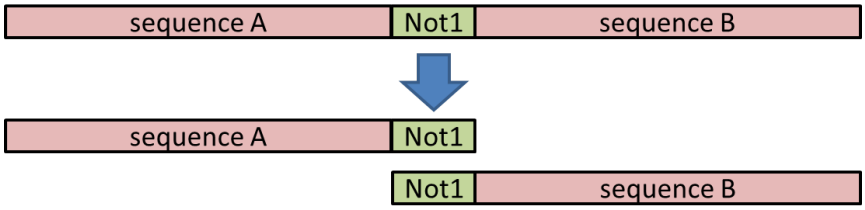
### 1. Generation of pooled DNA and sequencing of combined target exons

In total, 582 tumor DNA specimens from a variety of myeloid neoplasms (Supplementary Table 4) were analyzed for possible mutations within 9 genes, including *U2AF35, U2AF65, ZRSR2, SF1, SRSF1, SRSF2, SF3A1, SF3B1*, and *PRPF40B*. After genomic DNA was amplified using REPLI-g mini kit® (Qiagen), the concentration of the amplified genomic DNA was measured using the PicoGreen® reagents according to the manufacture's instruction. After adjusting the concentration of each amplified DNA to 75ng/μL, every 9μL of DNA from consecutive 12 samples were combined into one DNA pool, generating a total of 49 DNA pools. Next, the entire coding sequences of the 9 genes, which encompassed 39,376 bp in length, were amplified from each DNA pool by 120 independent genomic PCR with a NotI linker attached to each primer. After successful amplification was checked by gel electrophoresis, the products from the 120 PCRs for each DNA pool were combined together, followed by purification of DNA using QIAquick PCR Purification Kit®(QIAGEN) and digestion with NotI. The digested DNA was purified again and an aliquot of 2.5 μg of purified DNA was ligated with T4 DNA ligase for 5 hours, sonicated into ~200bp fragments on average using Covaris®, and used for generation of sequencing libraries, according to a modified Illumina pair-end library protocol. The libraries were then subjected to deep sequencing on Illumina GAIIx® according to the standard protocol for 108 bp pair end reads, where 8 DNA pools were analyzed in a single run on separate lanes.

### 2. Data analysis

#### 2.1 Preprocessing

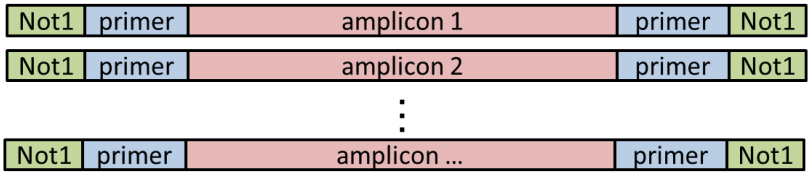
First, .fastq files originally generated from Illumina sequencers were converted to Sanger format files via the maq–sol2sanger software. After adaptor sequences were removed as described in the exon-sequencing analysis pipeline, each read was searched for the presence of NotI sites. When a NotI site was found within a read, the read was divided into two reads with that NotI site preserved (see below), where less than 10 bp reads were discarded.




#### 2.2 Mapping of sequences

After preprocessing, each read in the converted files was aligned to the set of the target sequences comprising the 120 target amplicons with their primer and NotI sequences

being preserved in both ends (see below), using BWA-0.5.8 with the –samse option to output the mapping results into .sam files, where the following reads are removed from the analysis, i.e., those having more than 4 mismatches to the reference, less than 30 of mapping quality, or soft clipped bases, to reduce the noise and artifact introduced during the alignment step.




**2.3 Detection of mutations**

The generated .sam files were converted to .pileup format via samtools mpileup command (http://samtools.sourceforge.net/). For each SNV or indel variant, which is defined as a mismatched base or a small nucleotide insertion or deletion, we calculated the mismatch rates, $R_i^+(v)$ and $R_i^-(v)$, at position (cycle number at which that variant was detected) $i$ in + and – strands, respectively, by enumerating the corresponding reads to be divided by the total number of read (see below table).

| | 1 | 2 | 3 | … | i | … | 108 |
|---|---|---|---|---|---|---|---|
| | $R_1^+(v)$ | $R_2^+(v)$ | $R_3^+(v)$ | … | $R_i^+(v)$ | … | $R_{108}^+(v)$ |
| | $R_1^-(v)$ | $R_2^-(v)$ | $R_3^-(v)$ | … | $R_i^-(v)$ | … | $R_{108}^-(v)$ |

The base calls in early and late cycles are prone to sequencing errors and some SNVs appeared in a strand dependent manner. To minimize the false positive calls due to these errors and to obtain a more robust estimation of the mismatch rates, we adopted the minimum of the median values of $\{R_i^+(v) \mid 11 \le R(v) < i \le 50\}$ and $\{R_i^-(v) \mid 11 \le i \le 50\}$ as the mismatch rate of that variant,. Next, synonymous variants, polymorphisms registered in the dbSNP131 and 1000 genome database, and variants on the intron region except splicing sites were filtered. Finally, the following multi-lane calling rule was performed, for removing the false positives that were caused via sequencing errors commonly occur among multiple lanes:

a) If $R(v) > 0.005$ for one lane and $R(v) = 0$ for the other lanes, then the lane with $R(v) > 0.005$ was investigated for the mutation $v$.
b) If $R(v) > 0$ for multiple lanes, then the lanes with $R(v) > 0.01$ were investigated for the mutation $v$.

**3. Validation of mutations and identification of the individuals carrying the mutations**

Each candidate mutation was validated by direct sequencing of the original individual amplicons from the 12 of the candidate pool.

## III. Preparation of high-titer virus stock

U2AF35 cDNA was purchased from TOYOBO co. (FCC137H10) and used as a PCR template for in vitro mutagenesis to generate the S34F, Q157P, and Q157R mutants, using QuikChange Site-Directed mutagenesis kit® (STRATAGENE). The wild-type and mutant U2AF35 mutants were constructed into the retrovirus vector pGCDNsamIRESEGFP[1] and tetracycline-inducible lentivirus-based expression vector, LV-TRE-PRE-Ubc-tTA-I2G[2].

For obtaining high-titer retrovirus stock, vector plasmids were co-transfected with a VSV-G expression construct into 293GP cells to recover retrovirus-containing supernatant, which was then used to transduce 293GPG cells (kindly provided by Dr R.C. Mulligan)[3] to establish stable cell lines capable of producing VSV-G-pseudotyped retroviral particles upon induction[4]. The average titer of retrovirus stocks prepared from these cell lines routinely exceeded approximately 1–10 × $10^7$ inclusion-forming units/mL, as estimated using Jurkat cells. Lentiviral vectors were generated by transient transfection of multiple plasmids into 293T cells as described previously[5].

## IV. Whole transcriptome analysis using GeneChip®Human genome U133 plus 2.0 Array and Exon 1.0 ST Array

### 1. Expression and Exon array experiment

Total RNA was extracted from HeLa cells retrovirally transduced with either mock, the wild-type *U2AF35* or the S34F mutant, and analyzed in triplicate using the GeneChip®Human Genome U133 plus 2.0 and Exon 1.0 ST array according to the manufacture's protocol.

### 2. Data analysis for Expression array

Raw array signals were processed using Expression Console software (Affymetrix), in which we adopted the Robust Multi-Chip Average (RMA)[6] method for background correction, normalization and summarization. GSEA v2.07 (http://www.broadinstitute.org/gsea/) and gene set collection of biological processes (c5.bp.v2.5.symbols; Gene Ontology) were used for following analysis. A gene set of NMD factors used in the analysis was constructed using 17 NMD core factors[4)].

### 3. Data analysis for Exon array

For data analysis, we used the Exon Array Analyzer, the web interface program which includes preprocessing of raw .CEL files, detection of alternative splicing and genome viewer (http://EAA.mpi-bn.mpg.de/supp.php)[7]. Raw array signals were extracted from .CEL files and background correction, normalization and summarization were performed using RMA. The standardized signal value was obtained for each probe set on each of triplicate array experiments, which were compared between three groups of experiments, i.e., those for mock, wild-type and mutant *U2AF35*-transcuced cells as follows:

First, the *p*-values were calculated by comparing the probe signals with the signals of the background probes using the Detection Above Background (DABG) method[8], and were used to statistically evaluate the significance of expression.

Second, a probe set or probe sets of a gene was excluded from the comparison to avoid too many false positive results[9],

1) when the probe set was not expressed in at least one of the three group, i.e., the DABG *p*-value was < 0.05 in less than 50% samples in both groups to be compared.
2) when the gene was not expressed in both groups, where a gene was considered to be expressed if more than 50% of probe sets had a DABG *p*-value < 0.05.
3) unless the probe set was assigned as "unique" in the Affymetrix annotation files, i.e., it has high potential for cross-hybridization.
4) when the gene showed more than 10-fold difference in the level of gene expression between the two groups.
5) when the probe set showed very large exon/gene intensity ratio. (Up to 5.0 was accepted.)

After excluding above probe sets, 375,526 probe sets were subjected to further analysis by calculating gene-level normalized intensities and the Splicing Index [10], which are defined as

$$\text{Genelevelnormalizedintensity} = \frac{\text{Probesetintensity}}{\text{Expressionlevel of the gene}}$$

$$\text{SplicingIndex(S.I.)} = \log_2 \frac{\text{S34Fnormalizedintensity}}{\text{WTnormalizedintensity}}$$

To obtain statistical evidence, the Student's t-test was performed on the gene-level normalized intensities and corrected for multiple comparison with the Benjamini & Hochberg method [11].

The exon array contains 1.4 million probe sets having different levels of evidence for representing true exons[12] (Supplementary Figure 13). The "Core" probe sets are supported by the most reliable evidence from RefSeq and full-length mRNA GenBank records. In addition to the Core set probes, the "Extended" probe sets contains those probe sets mapped to those sequences having other evidence of exons, such as ESTs, syntenic sequences to mRNA from rat and mouse, and micro RNA. The "Full" probe sets further includes putative exons only supported by computational gene prediction. Accordingly, the likelihood of representing true exons is expected to decrease for the Extended and the Full sets. In fact, the majority of the non-Core probe sets from the latter two are mapped to outside of the well-annotated exons and therefore, are likely to represent intron sequences rather than exons, even though they were not designed for the purpose of the analysis of intron sequences. To detect intron retentions due to abnormal RNA splicing, core probe sets as well as extended and full probe sets were analyzed.
To confirm the differential enrichment of Core and non-Core probe sets is not due to experimental error but significant inhomogenous distribution, we performed Mann-Whitney U test.
First, to evaluate experimental error, we compare two of three samples from the same cell (for example Mock No.1 vs Mock No.2, and there are 9 different pairs). Probe sets were ranked according to Splicing Index and U-value was calculated between Core and

non-Core probe sets. Second, we compare one of three samples from both wild-type and mutant *U2AF35*-transduced cells (for example wild-type No.1 vs S34F No.1, and there are 9 different pairs). U-value was calculated between Core and non-Core probe sets in the same way. Then, comparison of U-values of these two groups was performed using Wilcoxon test.

**V. Quantitative RT PCR**

We primed 500 ng of total RNA with 200 ng of random primers, which was subjected to reverse transcription using Superscript II® (Invitrogen) according to the manufacturer's protocol. The relative standard curve technique was used to analyze the expression of the gene of interest normalized to *GAPDH* expression in the same sample. The standard curves were prepared from equally pooled control cDNAs at 10, $10^2$, $10^3$ and $10^4$ dilution factors for each primer pair. Each reaction well contained 2 µl of template cDNA at appropriate concentrations for linear amplification based on the standard curve, 1.0-10 pmol of each primer and 1 x Brilliant III Ultra-Fast SYBR Green QPCR Master Mix® (Agilent Technologies) to a final volume of 20 µl. Reactions were carried out using a LightCycler 480 Real-time PCR System® for 50 cycles (95 °C for 10 seconds and 60 °C for 30 seconds). The purity of the PCR products was determined by melting curve analysis. Data were analyzed using LSC480 Software v.1.2.0.169 (Roche Applied Science). The primer set used for the analysis of NMD pathway genes are presented in Supplementary Table 10.

**VI. RNA sequencing**

**1. RNA sequencing procedure**

Total RNA from cultured cells was isolated with the RNeasy Total RNA kit (Qiagen) using RNase-free DNase set (Qiagen) to reduce contamination of genomic DNA according to the manufacturer's protocol. Libraries for sequencing were prepared using the Illumina mRNA-Seq Sample Preparation Kit, according to the manufacturer's instructions. Briefly, poly(A)+ RNA was recovered from 1µg of total RNA using oligo-dT - coated Sera-Mag magnetic beads. The recovered poly(A)+ RNA was then chemically fragmented. RNA fragments were converted to cDNA using SuperScript II and random primers. The second strand was synthesized using RNaseH and DNA Pol I. The ends of the cDNA were repaired using T4 DNA polymerase, T4 polynucleotide kinase, and Klenow DNA polymerase. A single adenosine was added to the 3′ end using Klenow fragment (3′ to 5′ exo minus). Adaptors were attached to the ends of the cDNA using T4 DNA ligase. 250 bp fragments were extracted from a 6% polyacrylamide gel electrophoresis (PAGE) gels (Bio-Rad). The 250 bp fragment was then amplified by 15 cycles of polymerase chain reaction (PCR) using Phusion DNA polymerase. Libraries were validated with an Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA). Libraries were diluted to 11 pM and applied to an Illumina flow cell using the Illumina Cluster Station. Sequencing was performed on an Illumina GAIIx with the pair end 108 bp read option, according to the manufacturer's instruction.

**2. data analysis**

For the analysis of the obtained sequencing data, the first 4bp and last 29bp bases were

removed from all reads in order to facilitate the alignment on short exonic regions and also to reduce the noise of sequencing errors. After the .fastq files were converted to the Sanger version of .fastq format with maq-sol2sanger program [http://maq.sourceforge.ned/], sequenced reads were aligned to the NCBI Human Reference Genome Build 37 with bowtie (version 0.12.7) [http://bowtie-bio.sourceforge.net/index.shtml] with -a --best --strata -m 1 -v3 -p8 option.

All the reads successfully mapped to the human genome were classified into the either of the following categories according to their positions on the genome, 'exons', 'introns', 'exon/intron or intron/exon junctions (E/I junctions)', and others, where the RefSeq database was used as the definition of exons and introns (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz). We defined an E/I junction as a 17 bp sequence, having 9 consecutive exon sequence followed by the contiguous 8 intronic or intergenic bases or vice versa. A read was classified as an E/I junction read, if one of E/I junction sequences was completely included in that read. On the other hand, a read was classified into either exons or introns, if the read did not belong to the E/I junctions and contained longer exon- or intron-derived than intron- or exon-derived sequences, respectively. The remaining reads were classified as intergenic sequences, since the bowtie will not map a read to two separate genomic regions. The result of the genome mapping is summarized below. The numbers of successfully mapped reads accounted for 54.8 % and 56.8 % of all reads in wild-type and mutant U2AF35-transduced cells.

| U2AF35 | | WT | | S34F | |
|---|---|---|---|---|---|
| **All reads** | | 184191858 | | 182700610 | |
| **All mapped reads** | | 100987737 | 54.827% | 103802395 | 56.816% |
| | Exons | 81879968 | 81.079% | 83324216 | 80.272% |
| | Introns | 7367729 | 7.296% | 8132337 | 7.834% |
| | E/I junctions | 817347 | 0.809% | 869283 | 0.837% |
| | Intergene | 10922693 | 10.816% | 11476559 | 11.056% |

The total length of the genomic regions covered by the successfully mapped reads was compared between the two experiments. In order to adjust the difference in the number of mapped reads, we randomly selected $10^8$ reads from all mapped reads from each experiment and calculated the number of bases in the reference genome covered by at least one read was calculated. We repeated the random selection 10 times and averaged the results.

The significance of the increase in the junction reads in mutant U2AF35-transduced cells was evaluated by testing the current observation against the null hypothesis H0: that the current set of E/I junctions was not real (or specially selected by 'God' or 'evolution'). Thus, we randomly simulate 10,000 sets of E/I junctions having the same number of junctions with the actual one (i.e. 429,478) from a total of 69,141,062 exonic positions, and for each simulated set of E/I junctions, calculated the odds ratio (OR) of junction reads between the two experiments, which is defined as

$$OR = \frac{\frac{\#\text{ of junction reads}^{mut}}{\#\text{ of total mapped reads}^{mut} - \#\text{ of junction reads}^{mut}}}{\frac{\#\text{ of junction reads}^{wt}}{\#\text{ of total mapped reads}^{wt} - \#\text{ of junction reads}^{wt}}}$$

, where superscripts, wt and mt, indicate the experiment for wild-type and mutant U2AF35 expressing cells.

We evaluated the relative expression of exogenous to endogenous U2AF35 by differentially enumerated exogenous from endogenous U2AF35 transcripts by matching each read to those sequences specific to either endogenous (AGCGTGACAGGAGGCGGTCGAGAGATCGTGAAAGATCTGGGCGATTCTGA ) or exogenous (ATCGTGAAAGATCTGGGCGATTCGACTACAAGGACGACGATGA CAAGTGA) U2AF35 transcripts, according to the absence or presence of the coding sequence for the FLAG tag.

**VII. Cell proliferation assays and assays for cell cycle and apoptosis analyses**

TF-1 cells and HeLa cells were engineered to express wild-type- or one of the mutant-U2AF35 by retrovirus or lentivirus-mediated gene transfer. For the assessment of gene expression, RNA samples were prepared using RNeasy® kit (Qiagen co.) from the cells sorted by GFP-marking using a MoFlo® FACS cell sorter (Beckman Coulter) 48 hrs after retoroviral tranduction, and subjected to subsequent analysis. For cell growth assays, the cells transduced with lentivirus vectors were inoculated into 96 well culture plates and the cell growth was monitored using Cell Counting Kit-8® (Dojindo co.) according to the manufacturer's protocol. Cell cycle analysis was performed by flow cytometry in accordance with standard protocols. HeLa cells were harvested after 48hrs doxycycline induction and fixed in ice-cold 70% (vol/vol) ethanol for overnight at 4°C. Before flow cytometry, fixed cells were added propidium iodide(PI) and RNase A and incubated at room temperature. DNA content was measured using a FACS Calibur flow cytometer (BD Biosciences, San Jose, CA) and analyzed using CellQuest software (BD Biosciences) and the FlowJo (TreeStar, Ashland, OR). Cell aggregates were gated out of the analysis based on the width of the PI fluorescence signal and each profile was obtained from 20,000 gated events. Apoptosis analysis was also performed using HeLa cells after 48hrs induction. Harvested cells were stained by APC-AnnexinV (BD Bioscience) and 7-AAD (Beckman Coulter) by manufacture's protocol and measured by FACS Calibur flow cytometer.

**VIII. Mouse KSL Competitive reconstitution assay**

**1. Purification of $CD34^{-}$KSL cells**

Mouse hematopoietic stem cells ($CD34^{neg/low}$, cKit-positive, Sca-1-positive, lineage-marker-negative: $CD34^{-}$KSL cells) were purified from bone marrow cells of 10-week-old male C57BL/6 (B6)-Ly5.1 mice according to the previously described protocol[13], with some modifications. The antibodies used for isolation of $CD34^{-}$KSL cells are listed below. Streptavidin-APC-Cy7 was also used (BioLegend).

| antigen | conjugation | clone | Supplier |
|---|---|---|---|
| CD34 | Alexa Flour 700 | RAM34 | eBioscience |
| c-Kit | PE-Cy7 | 2B8 | eBioscience |
| Sca-1 | Pacific Blue | D7 | BioLegend |
| CD4 | Biotin | L3T4 | eBioscience |
| CD8 | Biotin | 53-6.7 | eBioscience |
| Gr-1 | Biotin | RB6-8C5 | eBioscience |
| TER-119 | Biotin | TER-119 | eBioscience |
| B220 | Biotin | RA3-6B2 | eBioscience |
| IL-7R | Biotin | A7R34 | eBioscience |

**2. Transduction of $CD34^-$KSL cells**

$CD34^-$KSL cells were deposited into recombinant fibronectin fragment (CH296, Takara Bio)-coated 96-well microtiter plates at 300 cells per well, and were incubated in $\alpha$-MEM supplemented with 1% fetal calf serum, 100 ng/ml mouse stem cell factor (SCF), 100 ng/ml mouse thrombopoietin (TPO) and $10^{-4}$M $\beta$-mercaptoethanol for 24 hr. Cells were then transduced with a high-titer retrovirus vector at a multiplicity of infection of 600.

After transduction, cells were further incubated in S-Clone SF-O3 (Sanko Junyaku, Tokyo, Japan) supplemented with 0.1% BSA, 50 ng/ml SCF, and 50 ng/ml TPO. Transduction efficiency was determined by $GFP^+$ percentage 6 days later.

**3. Competitive reconstitution assay**

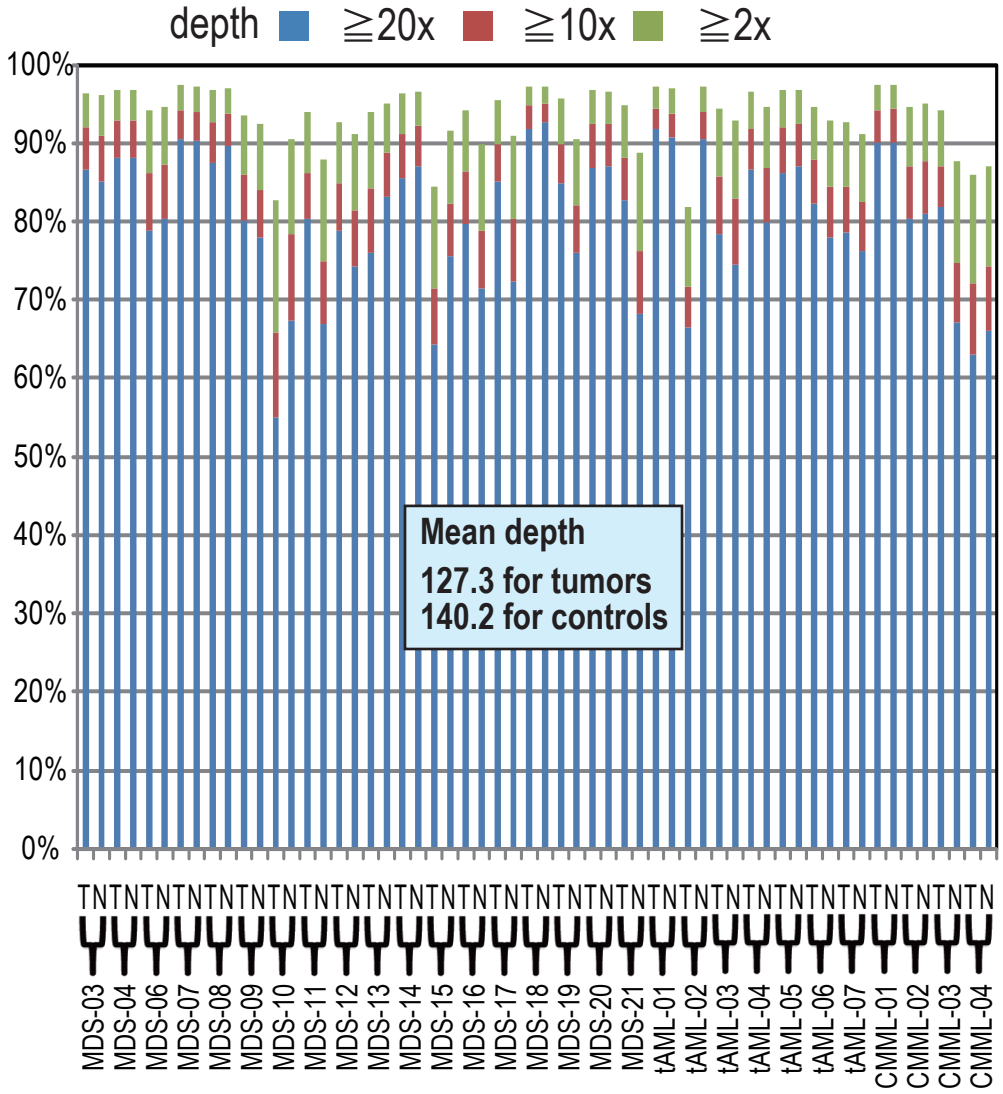
A competitive reconstitution assay was performed by transplanting transduced B6-Ly5.1 cells (expansion equivalent of 50 $CD34^-$KSL cells) together with 2 $x10^5$ whole bone marrow cells from 9-week-old B6-Ly5.1/5.2-F1 mice into each 9-week-old B6-Ly5.2 mouse lethally irradiated with two doses of 4.9 Gy, delivered 4 hr apart. Percent $GFP^+$ leukocytes were determined in recipient peripheral blood by flow cytometry analysis 6 weeks after transplantation. Animal experiments were performed with approval of the Institutional Animal Care and Use Committee of the Institute of Medical Science, University of Tokyo.


1 Nabekura, T., Otsu, M., Nagasawa, T., Nakauchi, H. & Onodera, M. Potent vaccine therapy with dendritic cells genetically modified by the gene-silencing-resistant retroviral vector GCDNsap. *Mol Ther* **13**, 301-309, doi:S1525-0016(05)01650-3 [pii]
10.1016/j.ymthe.2005.09.021 (2006).

2 Kobayashi, T. *et al.* Generation of rat pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. *Cell* **142**,

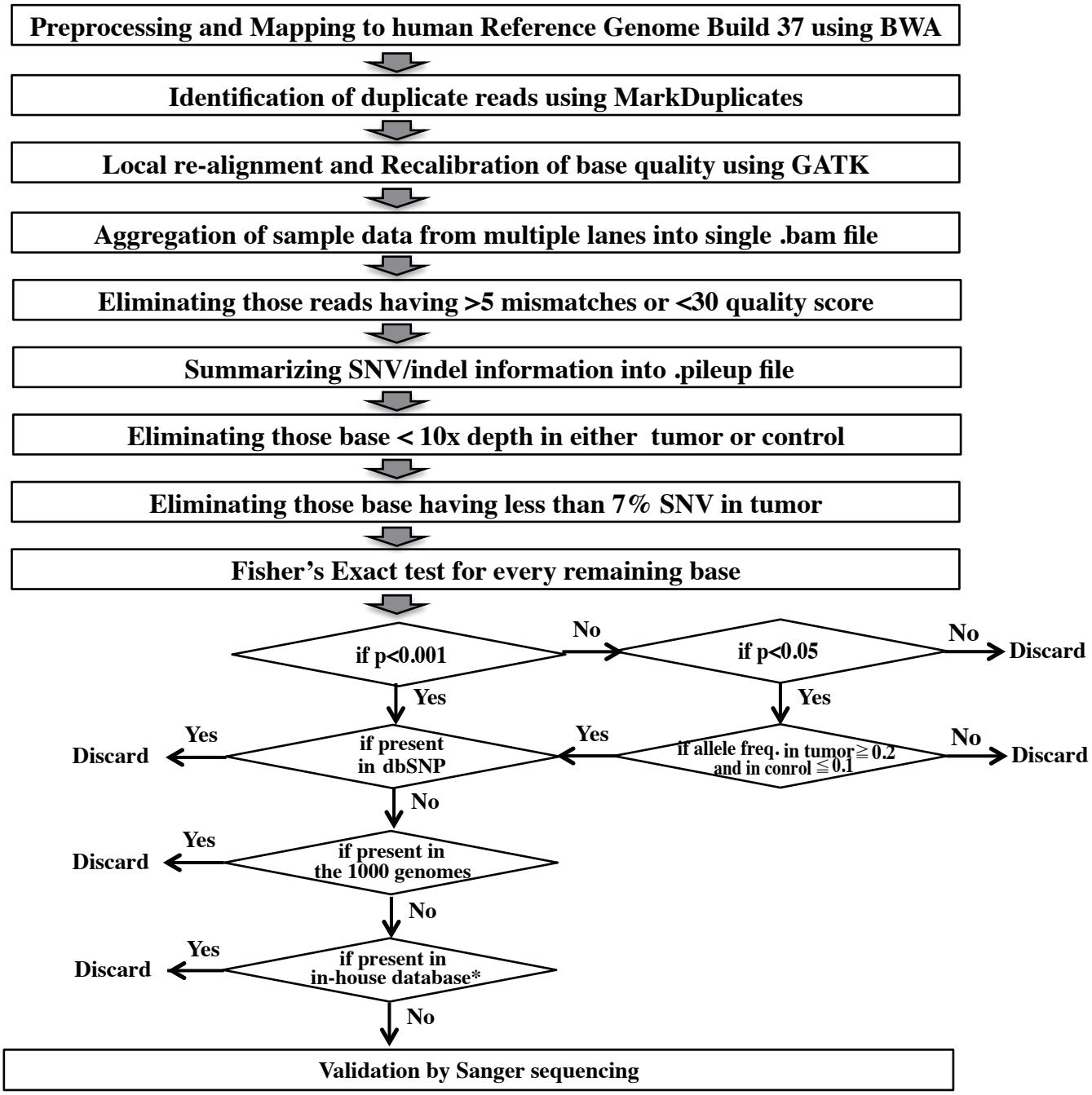
787-799, doi:S0092-8674(10)00843-3 [pii]
10.1016/j.cell.2010.07.039 (2010).
3 Ory, D. S., Neugeboren, B. A. & Mulligan, R. C. A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. *Proc Natl Acad Sci U S A* **93**, 11400-11406 (1996).
4 Suzuki, A. *et al.* Feasibility of ex vivo gene therapy for neurological disorders using the new retroviral vector GCDNsap packaged in the vesicular stomatitis virus G protein. *J Neurochem* **82**, 953-960, doi:1048 [pii] (2002).
5 Agarwal, S. *et al.* Isolation, characterization, and genetic complementation of a cellular mutant resistant to retroviral infection. *Proc Natl Acad Sci U S A* **103**, 15933-15938, doi:0602674103 [pii]
10.1073/pnas.0602674103 (2006).
6 Gellert, P., Uchida, S. & Braun, T. Exon Array Analyzer: a web interface for Affymetrix exon array analysis. *Bioinformatics* **25**, 3323-3324, doi:btp577 [pii]
10.1093/bioinformatics/btp577 (2009).
7 Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264, doi:10.1093/biostatistics/4.2.249
4/2/249 [pii] (2003).
8 Affymetrix, I. *White paper, Exon Array Background Correction v 1.0.*, (2005).
9 Affimetrix, I. *Technical Note: Identifying and Validating Alternative Splicing Events.*, (2006).
10 Clark, T. A., Sugnet, C. W. & Ares, M., Jr. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**, 907-910, doi:10.1126/science.1069415
296/5569/907 [pii] (2002).
11 Benjamini, Y. *Controlling the False Discovery Rate: A practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B, Statistical methodology.* (Royal Statistical Society, 1995).
12 Affymetrix, I. *White paper, Exon Probeset Annotations and Transcript Cluster Groupings v 1.0.*, (2005).
13 Ema, H. *et al.* Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nat Protoc* **1**, 2979-2987, doi:nprot.2006.447 [pii]
10.1038/nprot.2006.447 (2006).

# Supplementary Figure 1




**Mean coverage of whole-exome sequencing in 29 paired tumor/control samples**
The coverage of the target regions analyzed at indicated depth is plotted for each pair of tumor and normal specimens. The mean depth for all tumor and normal samples is also indicated.

# Supplementary Figure 2

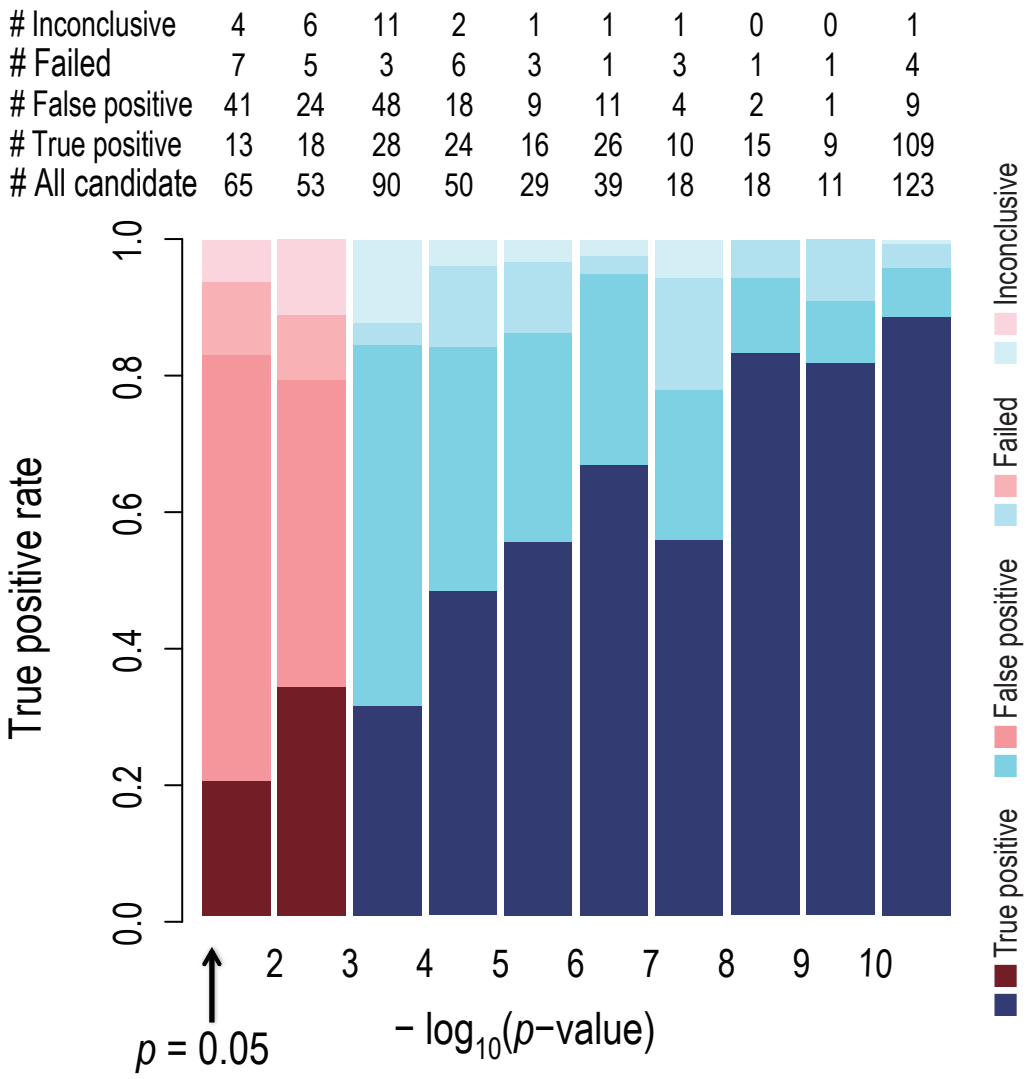



**The data processing pipeline for detecting somatic nonsynonymous mutations in whole-exome sequencing operating on the PC clusters at the Human Genome Center, Institute of Medical Science, University of Tokyo.**

Raw sequencing data is first summarized for each nucleotide base in both tumor and reference samples, followed by detection of candidate somatic mutations through evaluating the significance of difference in allele frequencies between tumor and germline control as well as the absolute allele frequencies in tumor sample. The former provides robust detection of mutations in the face of the presence of substantial contamination of germline samples (CD3+ cells or buccal mucosa).

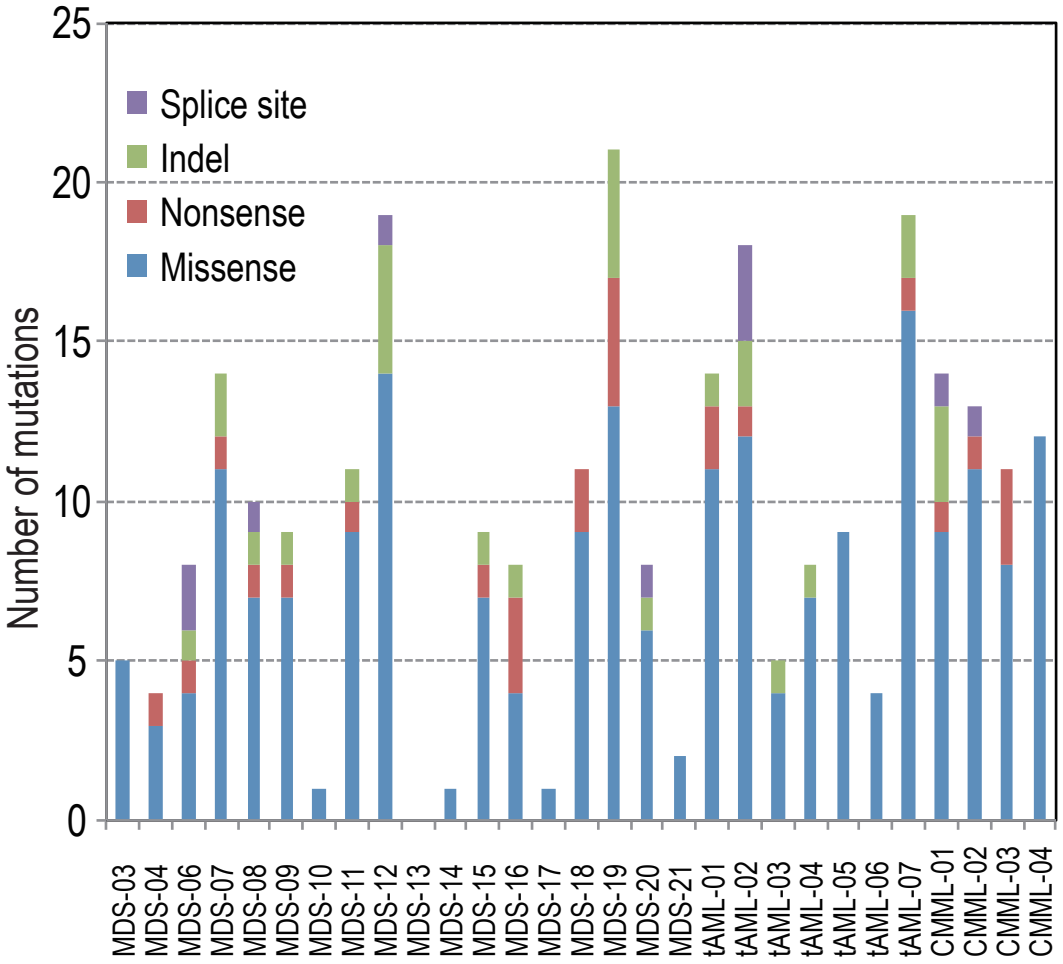
*The in-house database consists of SNVs and Indels detected in whole exome analysis of control samples in our laboratory.

# Supplementary Figure 3




**True positive rates in Sanger-based validation of candidate somatic mutations**
To see the performance of our pipeline for the identification of somatic mutations, true positive rate (TPR) was calculated for each bin defined by the logarithm of the Fisher's p-value, since our algorithm for mutation call primarily depends on the Fisher's exact test regarding the fraction of reference and candidate mutant alleles. The validated and non-validated changes are shown in thick and thin colors, respectively. As the p-value increases to the point where the effect of multiple testing cannot be neglected (0.001), the number of candidate SNV or indels to be validated increases, while the TPR dramatically decreases. To catch more mutations while maintaining the efficiency of validation within an acceptable range, we limit the candidates for further validation to those alleles whose frequency is >0.20 in tumor samples. With this approach, the TPR transiently recovers for $0.01>p>0.001$, although it decreases again for $0.5>p>0.01$. For higher $p$-values, the efficiency of validation becomes unacceptably low, which provides a basis for the threshold in our algorithm.

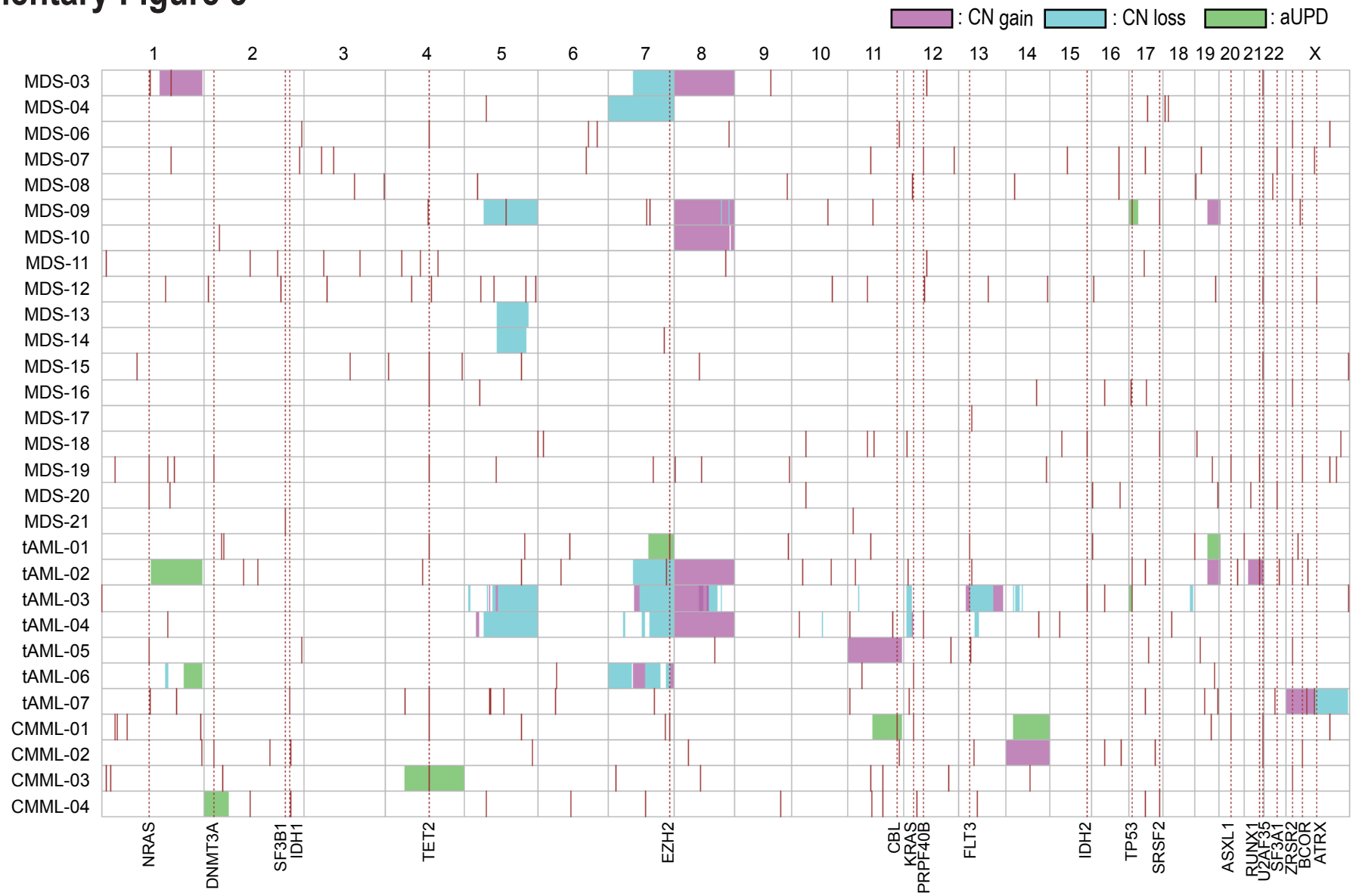
# Supplementary Figure 4




**Number of somatic changes in 29 patients with myelodysplasia**

The number of somatic mutations and indels are displayed for each patients with myelodysplasia. The type of somatic changes are discriminated by indicated colors.

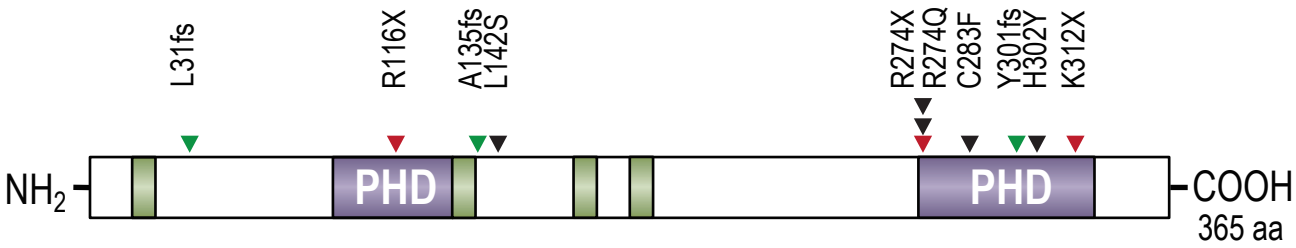
# Supplementary Figure 5




**Landscape of myelodysplasia genomes as revealed by whole-exome analysis combined with SNP array karyotyping**

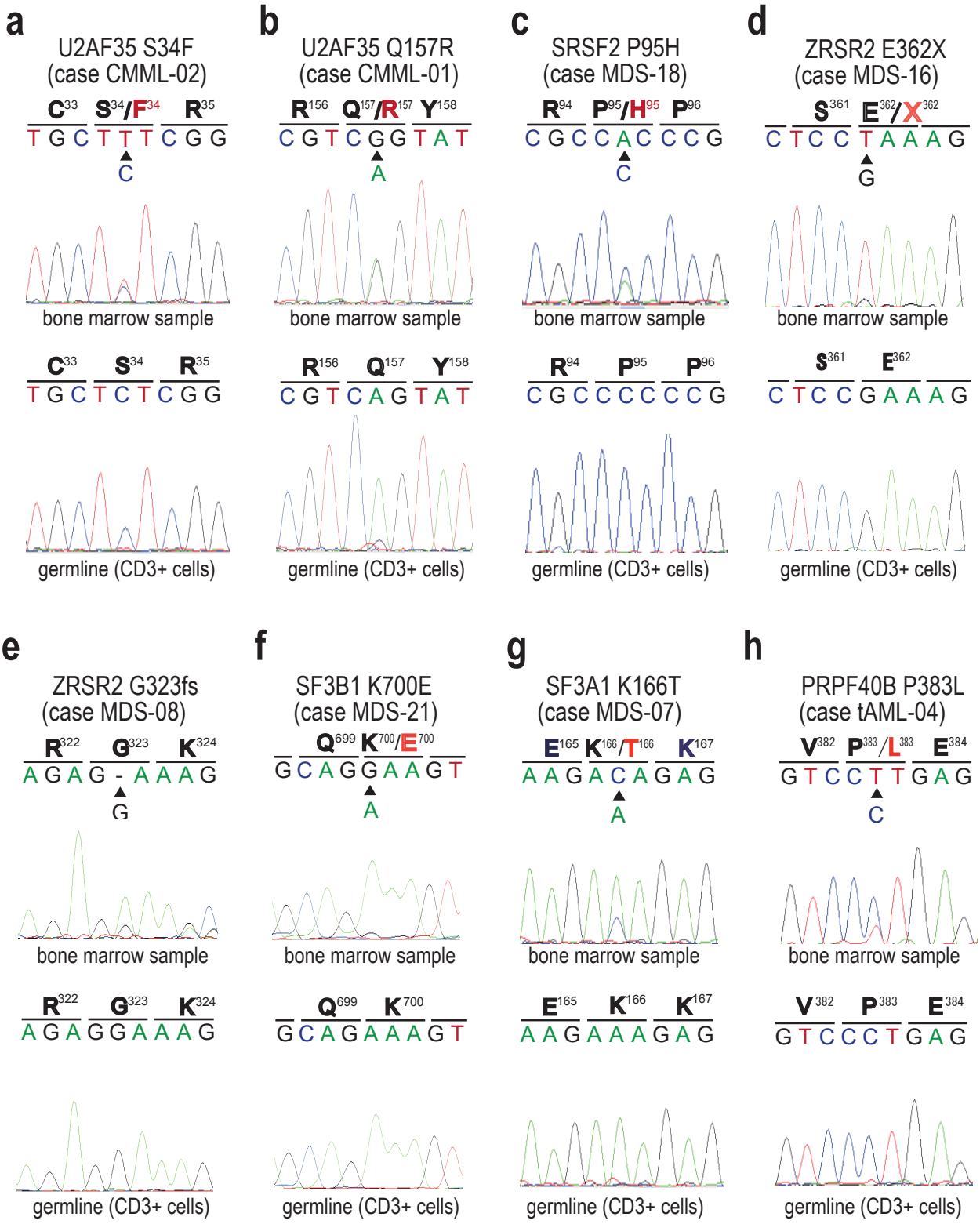
Summary of genetic analysis of 29 cases with myelodysplasia, in which somatic mutations found in whole exome-sequencing and copy number abnormalities detected by SNP array karyotyping using Affymetrix GeneChip 250K Nsp arrays are integrated into a whole genome view. Copy number abnormalities are shown in colored boxes as indicated. A total of 268 somatic mutations are shown in vertical lines, where the locations of recurrent mutations and spliceosome pathway gene mutations, as well as, previously known gene targets are indicated by broken lines.

# Supplementary Figure 6




***PHF6* mutations in MDS and related myeloid neoplasms**
Eleven mutations of *PHF6* found in 10 out of 164 (6.1%) cases are shown in arrowheads in red (nonsense), green (frameshift), and black (missense).

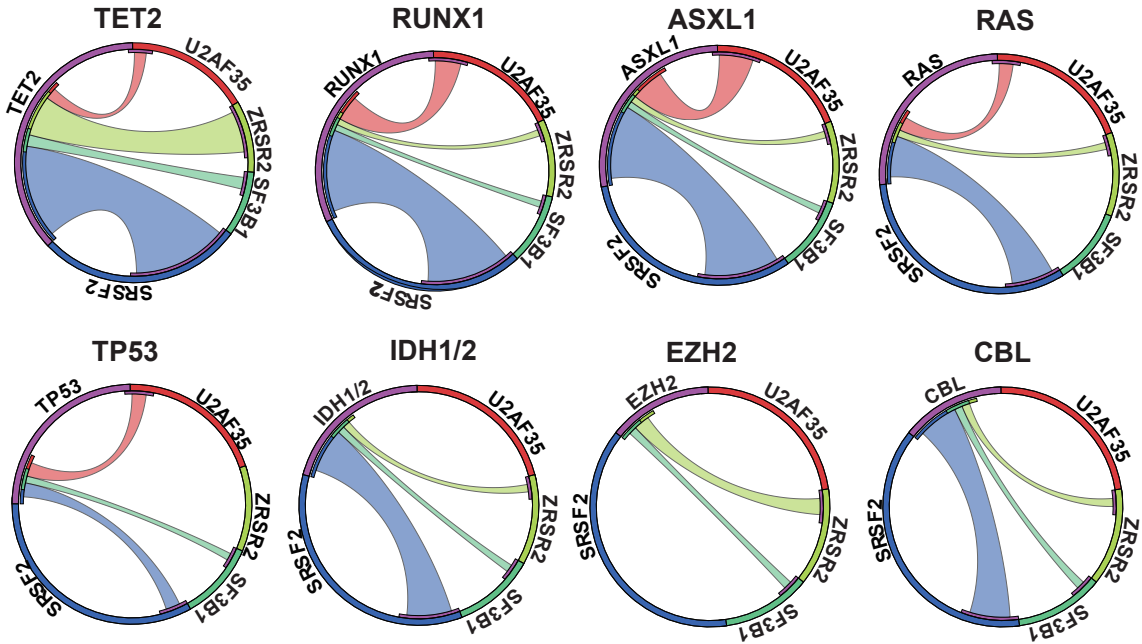
# Supplementary Figure 7




**Validation of somatic mutations in 6 spliceosome pathway genes detected by whole-exome sequencing by Sanger sequencing**

All 16 somatic mutations detected in 29 cases with myelodysplasia were validated by Sanger sequencing of both tumor and germline controls, of which representative results of 8 mutations found in 6 genes are presented (**a~h**). In each panel, sequences in tumor (top) and CD3+ T cells (bottom) are shown. The mutated nucleotides are indicated by arrowhead together with those in germline samples. Altered amino acids are indicated by red.

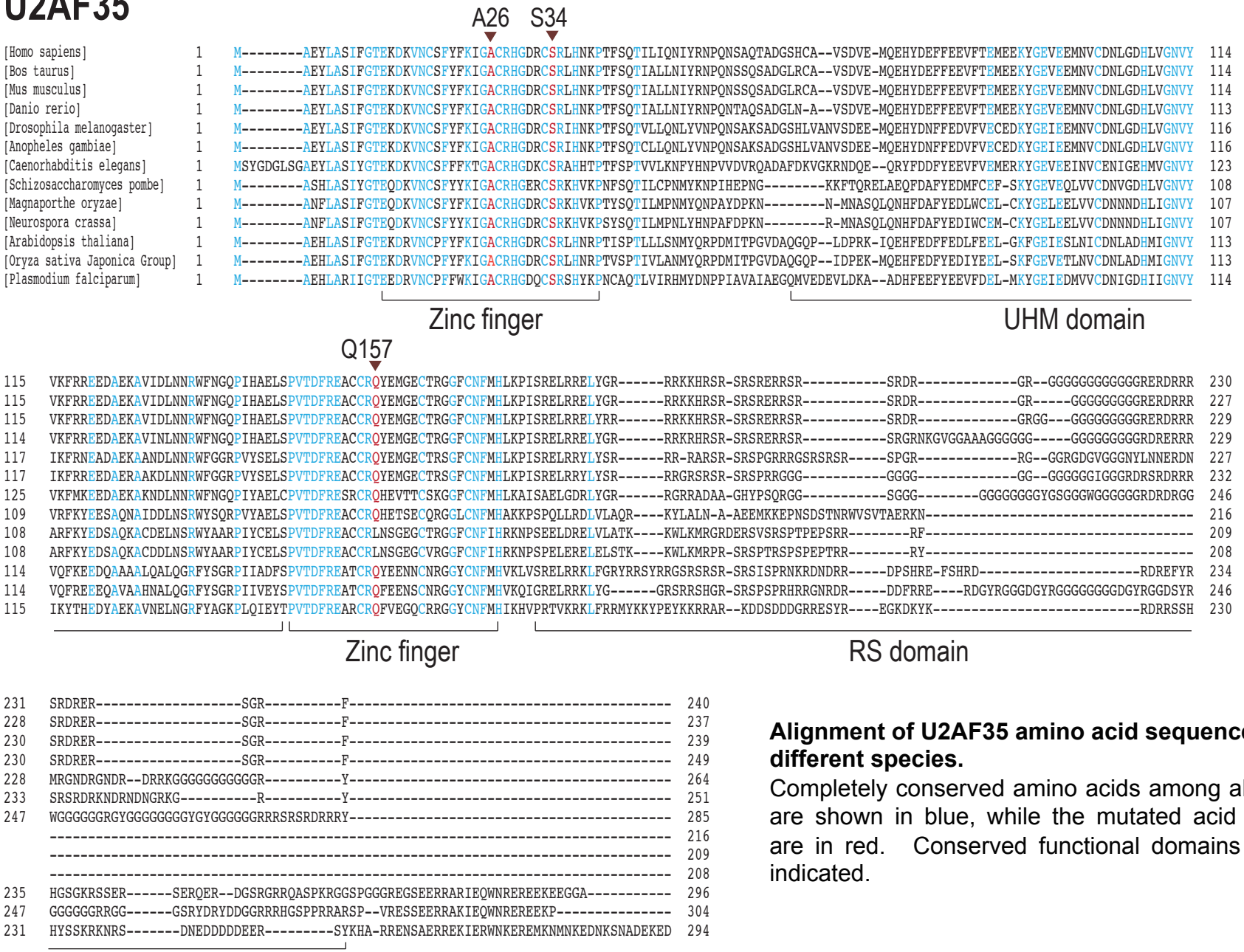
# Supplementary Figure 8




**Overlapping of spliceosome pathway gene mutations with common gene targets in myelodysplasia**

Overlapping of spliceosome pathway mutations with previously reported gene targets in myelodysplasia, including *TET2, RUNX1, ASXL1, RAS (NRAS* and *KRAS*), *TP53, IDH1/2, EZH2,* and *c-CBL* is shown for each of the latter genes. The mutations of common gene targets were analyzed in 161 out of the 582 cases with myelodysplasia (unpublished data). The arch length is proportional to the number of indicated mutations.

# Supplementary Figure 9

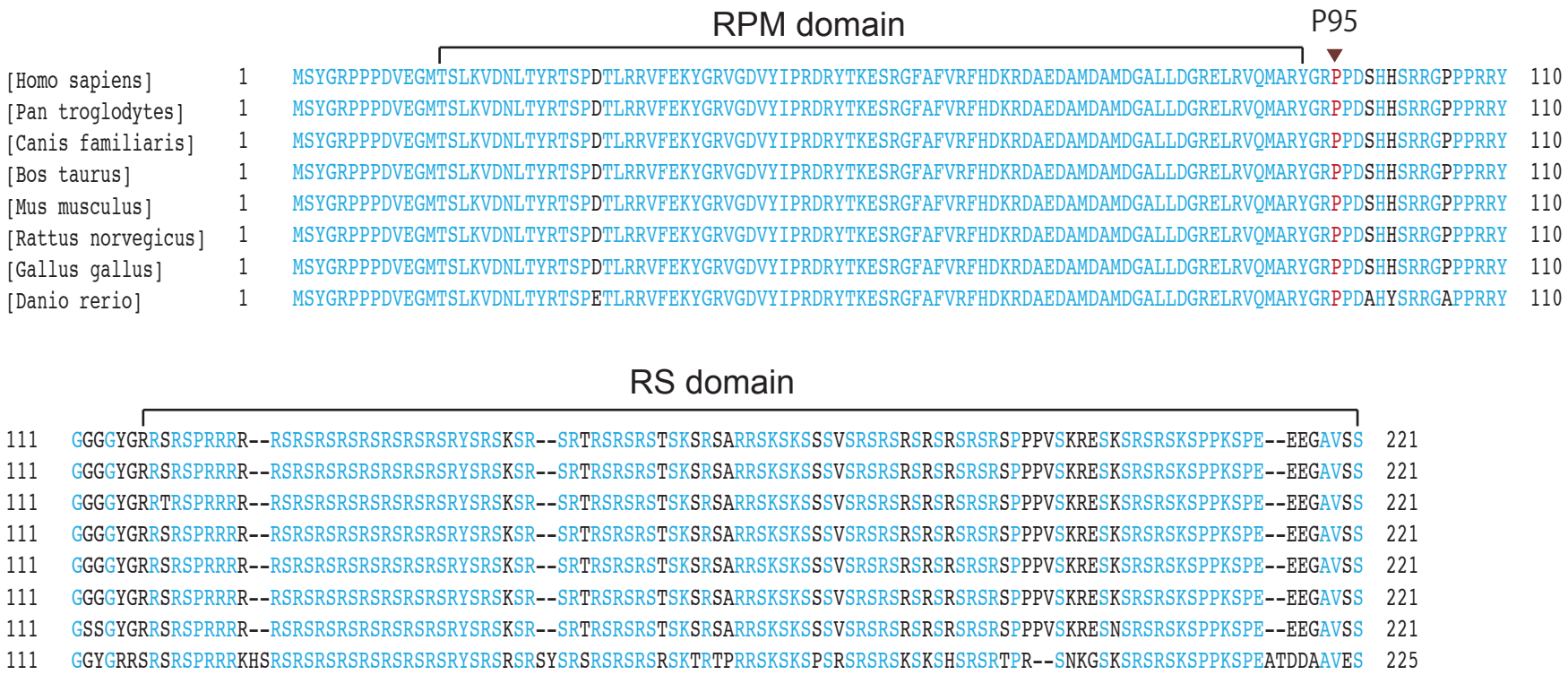



**Alignment of U2AF35 amino acid sequences of different species.**

Completely conserved amino acids among all species are shown in blue, while the mutated acid positions are in red. Conserved functional domains are also indicated.

# Supplementary Figure 10

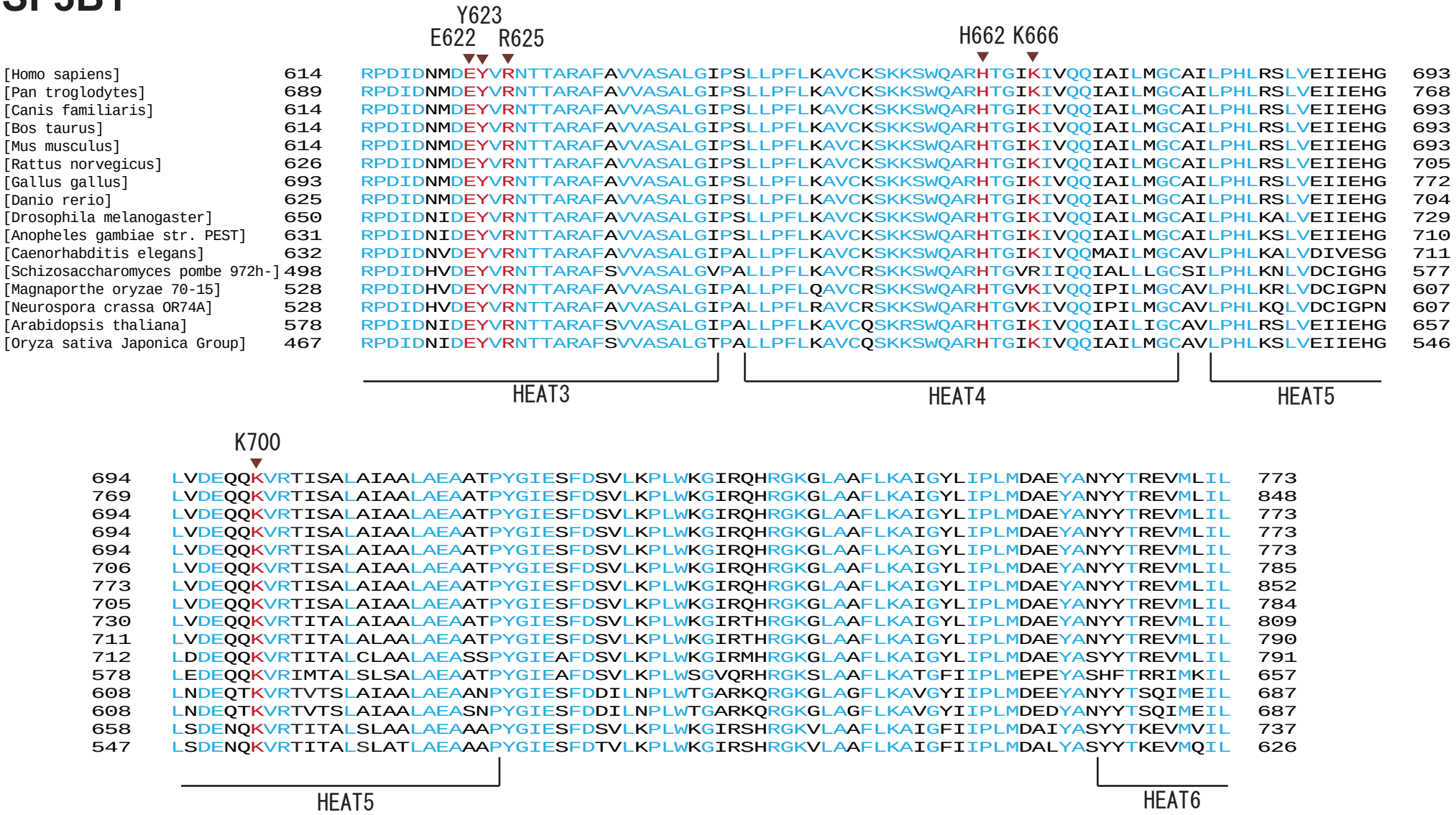
## SRSF2




**Alignment of SRSF2 amino acid sequences of different species.**
Completely conserved amino acids among all species are shown in blue, while the mutated acid positions are in red. Conserved functional domains are also indicated.

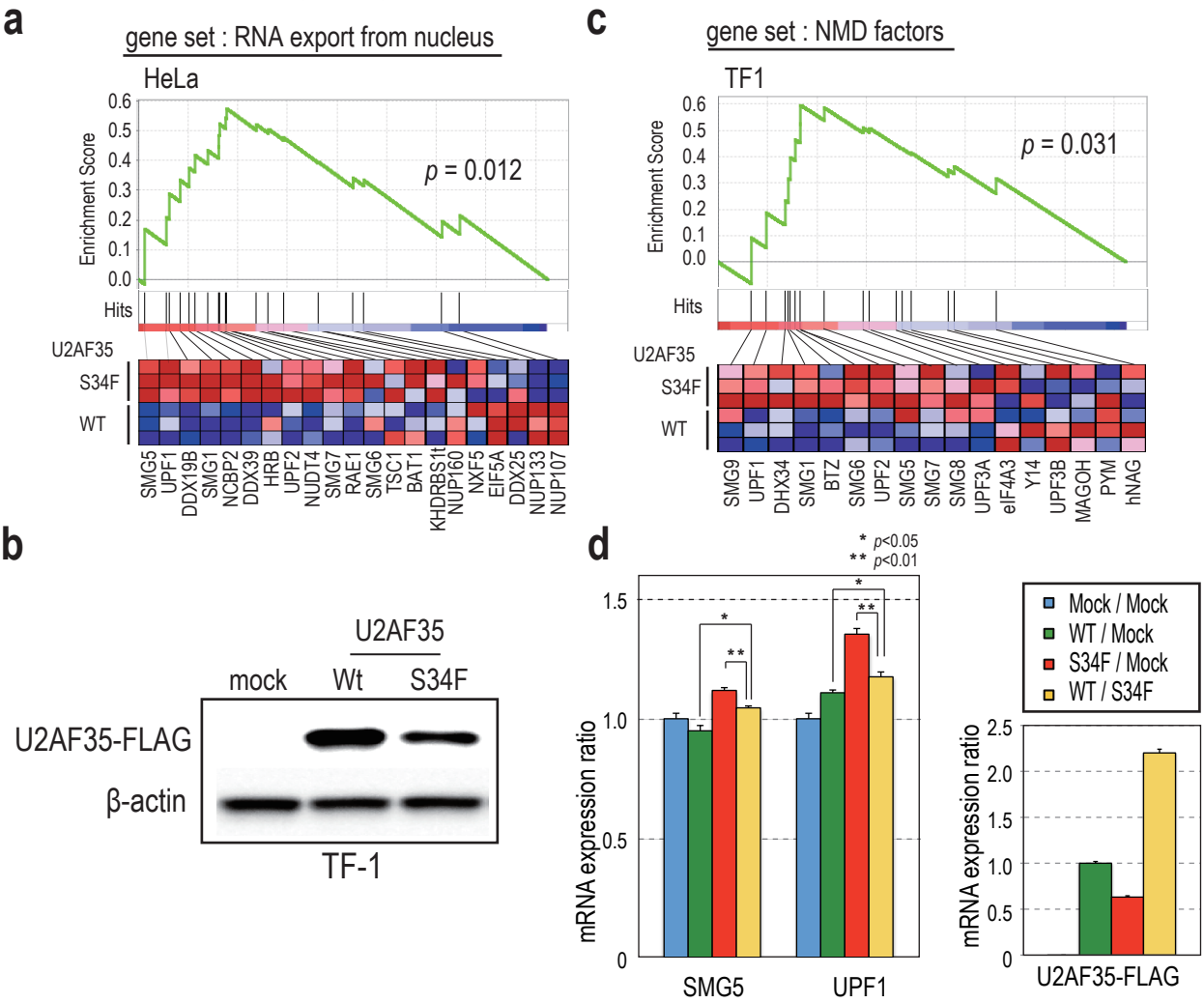
# Supplementary Figure 11

## SF3B1




**Alignment of SF3B1 amino acid sequences of different species.**
Completely conserved amino acids among all species are shown in blue, while the mutated acid positions are in red. Conserved functional domains are also indicated.

# Supplementary Figure 12

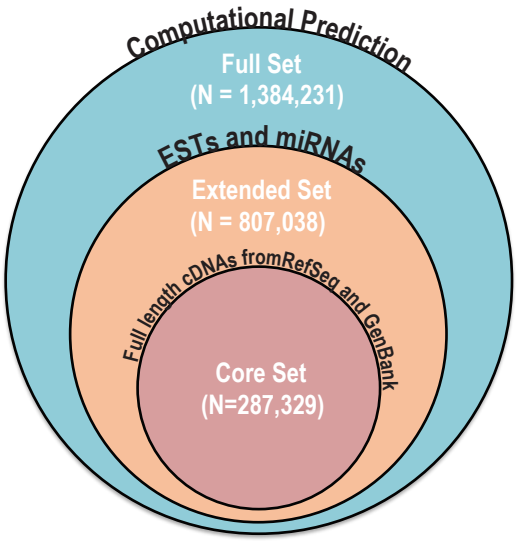



**Activation of the NMD pathway in HeLa and TF-1 cells transduced with either wild-type and mutant U2AF35**

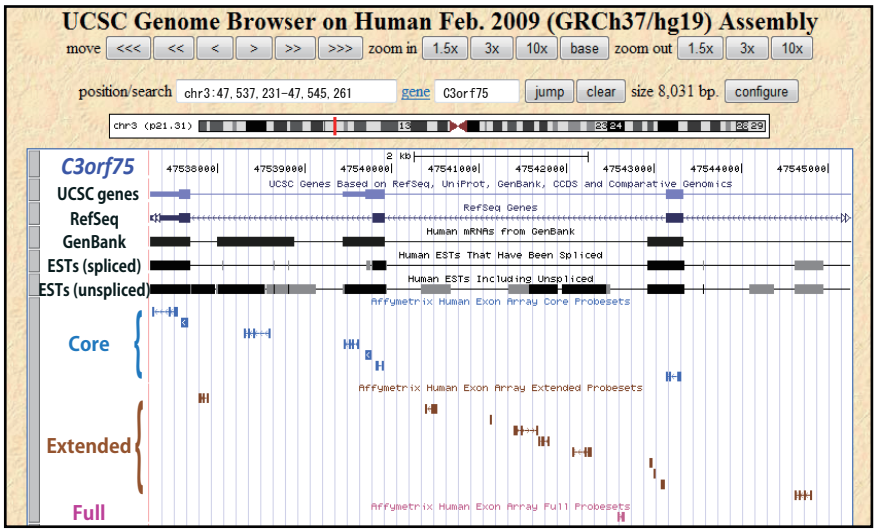
**a.** The result of the initial GSEA using gene set collection of biological processes (c5.bp.v2.5.symbols; Gene ontology) showing significant enrichment of the gene set named RNA export from nucleus in mutant *U2AF35*-transduced HeLa cells compared to wild-type *U2AF35*-transduced cells. In the next step, we selected the genes involved in NMD from the RNA export from nucleus gene set and constructed a more complete set for the NMD pathway genes with additional NMD-related genes. The significance of gene sets was empirically determined by 1,000 gene-set permutations. **b.** Western blot analyses showing expression of transduced wild-type or mutant (S34F) *U2AF35* in TF-1 cells used for gene expression analyses. **c.** The result of the GSEA of mutant *U2AF35*-transduced TF-1 cells also showing the enrichment of the NMD pathway gene set. The heat map of the expression of the target genes in this set is depicted at the bottom in each panel. **d.** Expression level of NMD related genes in wild-type and mutant *U2AF35* co-transduced HeLa cells. Wild-type and mutant *U2AF35* expression vectors (pGCDNsamIRES-EGFP or -kusabira orange) were co-transduced in HeLa cells by the calcium phosphate transfection method. Transfected cells were harvested 72 hours after transfection. RT-qPCR were performed to evaluate the expression level of NMD related genes. Mean and S.E. from three experiments of each NMD gene expression was plotted after normalization using GAPDH expression level in the same sample (left panel). Expression of transduced *U2AF35* in each sample was confirmed using RT-qPCR of U2AF35-FLAG (right panel). *P*-values were determined by t-test. U2AF35-FLAG-Forward primer: CCG CCG TCG CAA GAA GCA TA, -Reverse primer: CAC TTG TCA TCG TCG TCC TTG T.

# Supplementary Figure 13

**a**

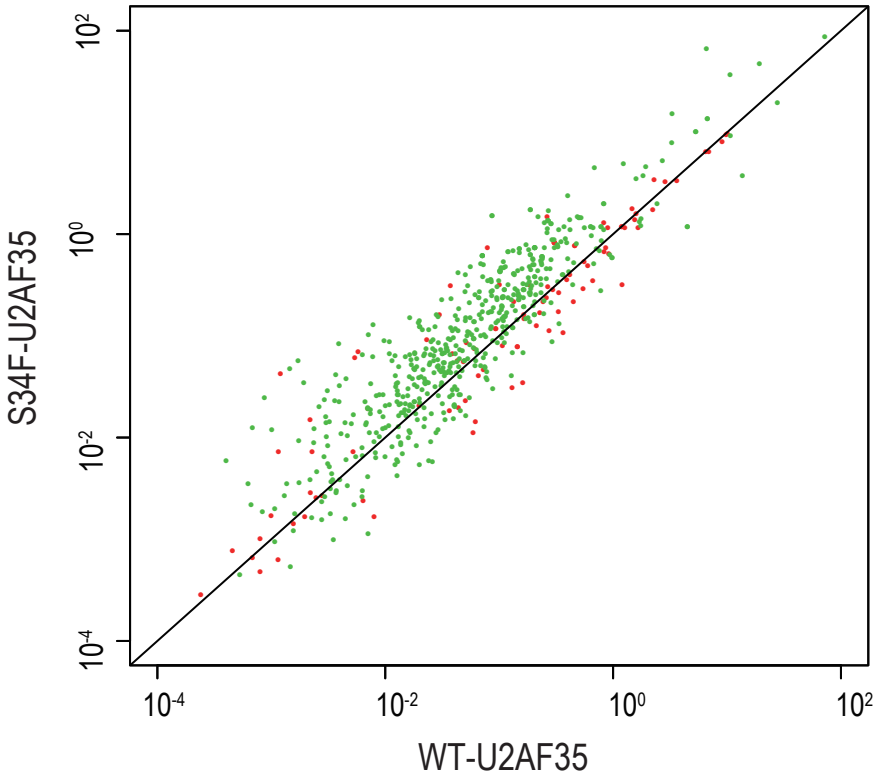



**b**



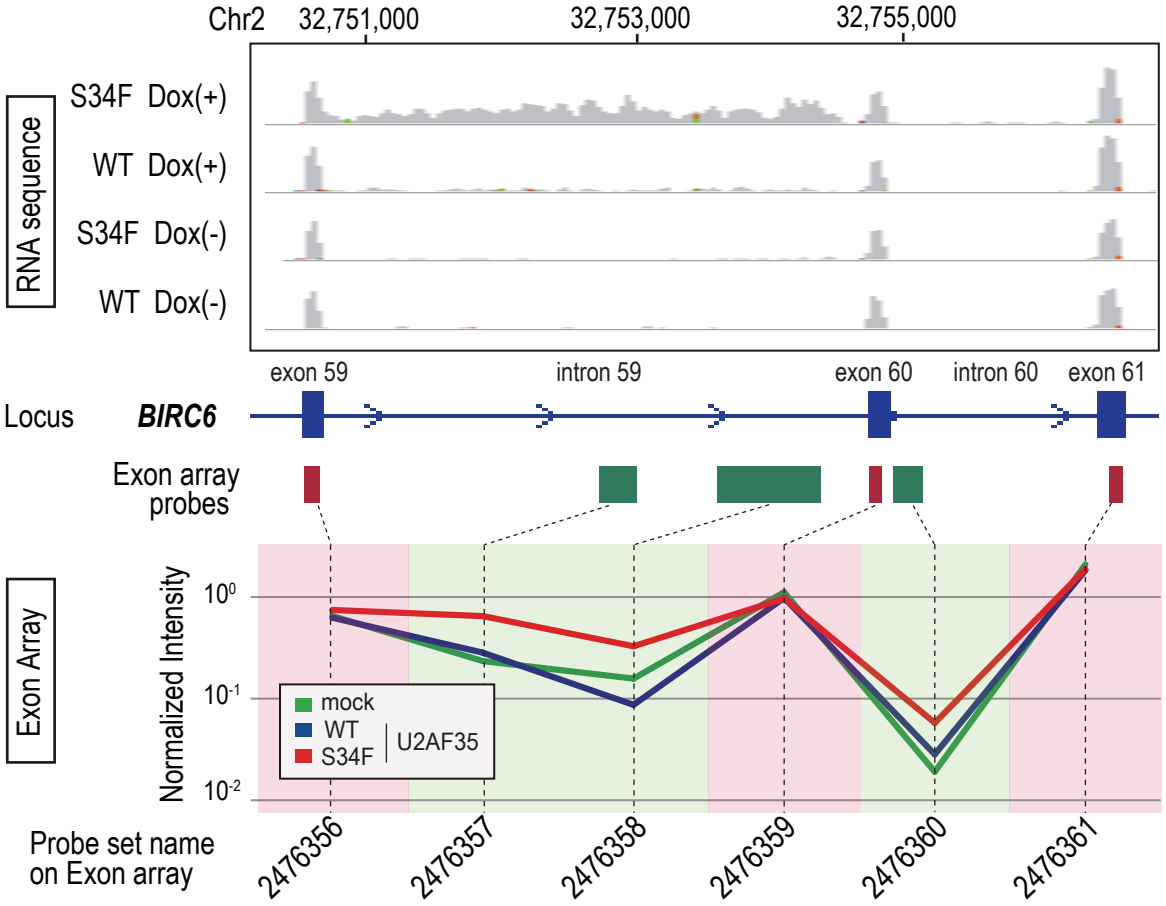

**The three discrete sets of probes on GeneChip® Human Exon 1.0 ST Array**
**a.** GeneChip® Human Exon 1.0 contains a total of 1.4 millions of probe which are divided into three categories. The Core set consists of ~290,000 probe sets derived from the full cDNA sequences registered in RefSeq or GenBank, consisting of highly reliable exon sequences. The extended set contains additional ~520,000 probe sets generated based on sequences from other ESTs and miRNAs. The full set further includes other ~580,000 probe sets, having computationally predicted exon sequences. The sequences of the probe sets in the latter two categories are much less reliable for exons. **b.** Typical examples of probe sets from the three different categories, where most probe sets in the Extended and Full sets are actually thought to be intron sequences.

# Supplementary Figure 14




**Differential enrichiment of exon and intron reads reproducibly observed in RNA sequencing**
Pair-wise scatter plots of normalized exon/intron reads in RNA sequencing. Number of exon/intron reads per $10^8$ total mapped reads were divided by gene expression level (FPKM-value; fragments per kilobase of exon per million mapped reads). Red and green plots indicate exon and intron respectively, which contain significantly differentially expressed probe sets between wild-type and mutant U2AF35-transduced cells in exon array.

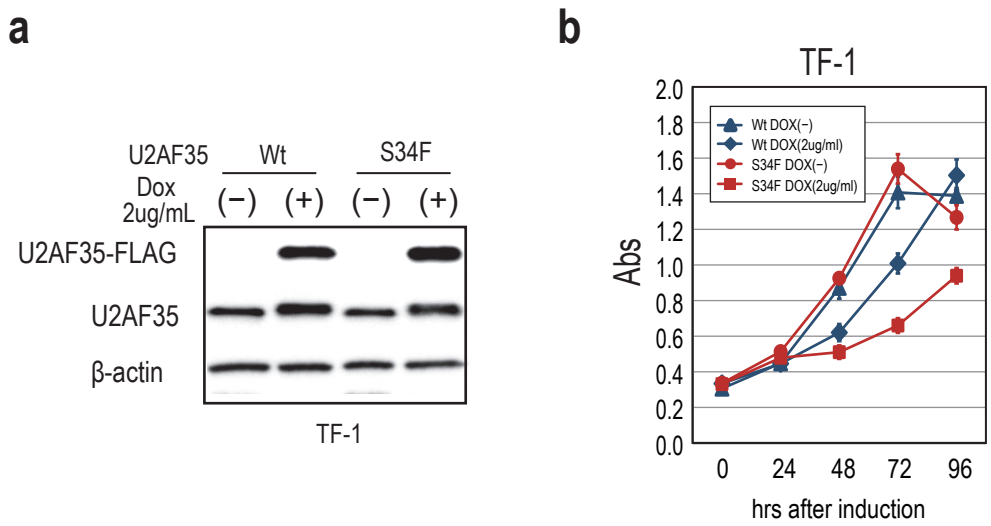
# Supplementary Figure 15




**A typical example of abnormal splicing caused by mutant U2AF35 detected by exon arrays and RNA sequencing**

On exon arrays, unspliced intron can be detected when one or more probe sets are placed in the intron. Many non-Core probe sets actually represent not in true exons but introns. For example, as shown in the middle panel, three probe sets within the intron 59 and 60 of the *BIRC6* gene indicated in green bars are included in the exon array together with three probe sets for the three authentic exons in RefSeq. Reflecting the fact that the intron is not spliced in the mutant U2AF35-transduced HeLa cells, the array signals from these three intronic probe sets increased compared to the signals from wild-type U2AF35-transduced cells or mock, while the three exonic probe sets showed comparable levels of signals among the three samples (bottom panel), indicating that the intervening intron sequences were not spliced out in the mutant-transduced cells. On the other hand, the high-throughput sequencing of RNA species in these cells reveals this splicing alterations more directly. As shown in the top panel, the abnormally spliced intron sequences can be explicitly detected by enumerating those reads mapped to the intron. In addition, the unspliced intron can be directly confirmed by detecting the reads that encompass the exon/intron boarders. Through the analysis of these exon/intron structures using exon arrays and/or RNA sequencing, we demonstrated that the mutant U2AF35 protein induced extensive abnormal splicing in HeLa cells as presented in Figure 3 g-h. This gene (*BIRC6*) is known to have anti-apoptotic function and therefore, is one of the candidate targets of the U2AF35 mutant. Loss of function of *BIRC6* induced by the aberrant splicing could be responsible for the enhanced apoptosis found in the mutant U2AF35-transduced cells.

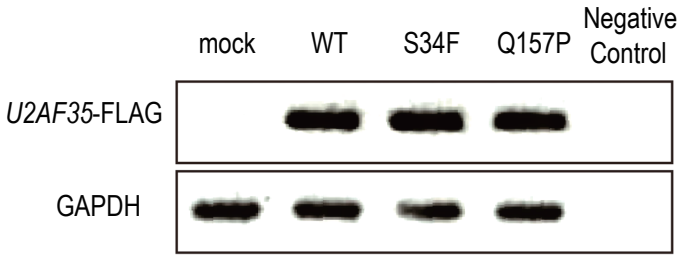
# Supplementary Figure 16




**Suppression of cell growth of TF-1 cell line transduced with mutant U2AF35**

**a.** Western blot analysis of TF-1 cells transduced with doxycycline-inducible U2AF35 constructs. Expression of U2AF35 protein was detected by anti-FLAG antibody and anti-U2AF35 antibody after the induction of expression by doxycycline. Anti-beta actin blot was used as lording control. **b.** Cell proliferation assays for TF-1 cells, showing significant growth suppression after induction of U2AF35 expression in mutant *U2AF35*-transduced cells but not in wild-type *U2AF35*-transduced cells. Cell numbers were measured using Cell-counting kit and were plotted as their mean absorbance ±S.D. (n = 6 ).

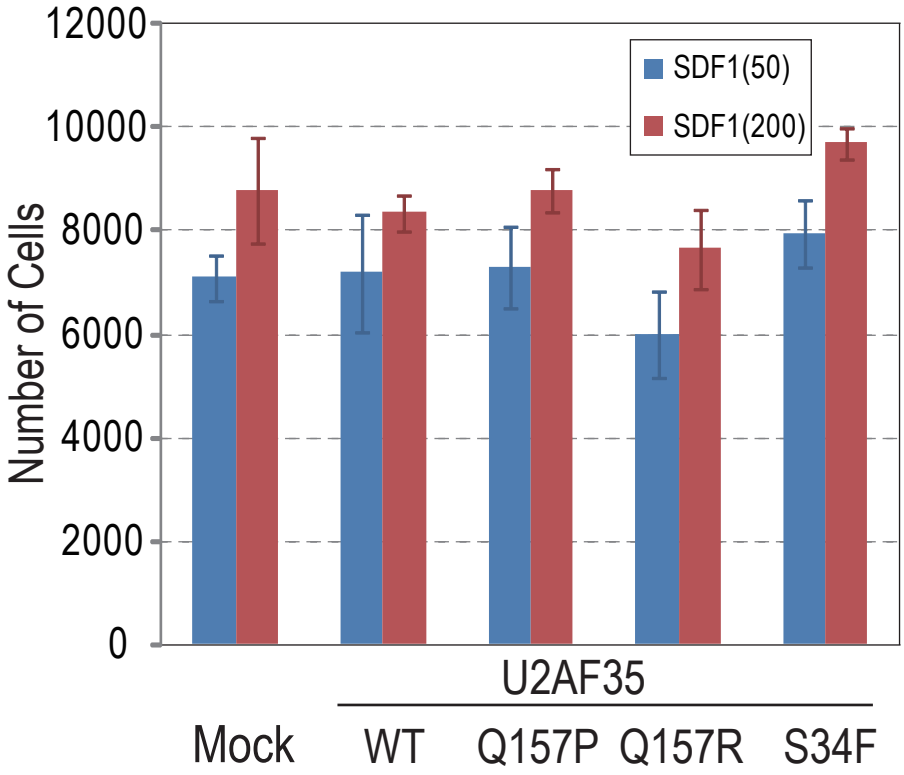
# Supplementary Figure 17




**Expression of transduced U2AF35 in mouse CD34$^-$ KSL cells**

RNA samples were prepared using RNeasy kit (Qiagen). RT-PCR of U2AF35-FLAG was performed for confirmation of the expression of transduced U2AF35 in mouse CD34$^-$ KSL cells. GAPDH-Forward primer:TGA CGT GCC GCC TGG AGA AA, -Reverse primer: AGT GTA GCC CAA GAT GCC CTT CAG.

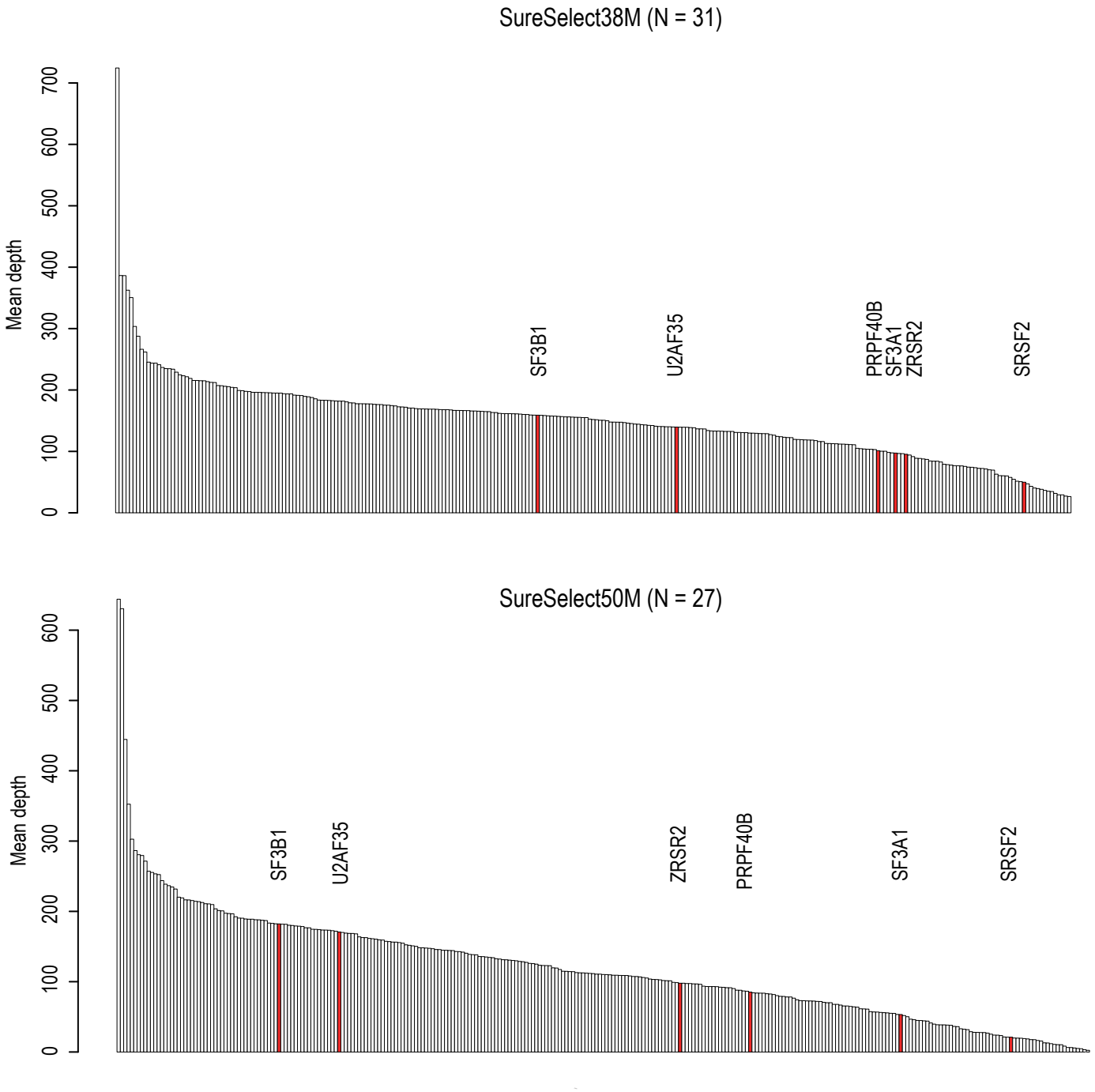
# Supplementary Figure 18




**Transwell® migration assays of mouse progenitor cellstransduced with U2AF35 mutants**

KSL cells were harvested from B57BL/6 mice and retrovirally trasnduced with indicated U2AF35 and mock constructs with EGFP marking. After transduction, a total of 1.4 x $10^4$ GFP positive cells were placed into the upper chamber of 5.0um HTS Transwell (Corning, NY) with the lower chamber containing 50 or 200 ng/mL recombinant murine SDF-1a/CXCL12 (Peprotech). After 4 hours incubation, the number of cells migrated to the lower chamber was counted by flowcytometry using Flow-Count Fluorospheres (Beckman Coulter). There was no significant difference in the number of cells that migrated into the lower chamber among different U2AF35-transduced cells at either SDF concentrations.

# Supplementary Figure 19




**The mean depth of analysis for genes involved in RNA splicing**

The read depth across all samples analyzed on 38M (N = 31) and 50M (N = 27) platforms was averaged over the entire coding sequence for each of 275 (in 38M platform) or 291 (in 50M platform) genes involved in RNA splicing, and plotted separately for each platform. The 6 genes ( *SF3B1*, *U2AF35*, *SRSF2*, *ZRSR2*, *PRPF40B* and *SF3A1* ) in which somatic mutations were detected in whole exome studies are indicated in red. More than ×10 depth was obtained in 100 % or 92.1 % of genes included in 38M or 50M platform.

**Supplementary Table 1. Characteristics of 29 patients subjected to whole-exome analysis**

| Sample ID | Sex | Age | WHO Classification | Cytogenetics | IPSS score | Target | Sample(Tumor) | Germline control |
|---|---|---|---|---|---|---|---|---|
| MDS-03 | M | 45 | RA | 47,XY,+1,der(1;7)(q10;p10),+8 | 0 | 38MB | Bone marrow | CD3+ T cell |
| MDS-04 | M | 72 | RA | 46,XY | 0 | 38MB | Bone marrow | CD3+ T cell |
| MDS-06 | M | 73 | RCMD | 46,XY | ND | 38MB | Bone marrow | CD3+ T cell |
| MDS-07 | M | 70 | RCMD | 47,XY,+8 | 1 | 38MB | Bone marrow | CD3+ T cell |
| MDS-08 | M | 75 | RCMD | 46,XY | 0.5 | 38MB | Bone marrow | CD3+ T cell |
| MDS-09 | M | 86 | RCMD | 46,XY,add(5)(q11.2),add(8)(q24.1) | ND | 50MB | Bone marrow | CD3+ T cell |
| MDS-10 | F | 25 | RCMD | 47,XX,+8 | 1 | 50MB | Bone marrow | Buccal mucosa |
| MDS-11 | M | 64 | RCMD | 46,XY | 0.5 | 50MB | Bone marrow | CD3+ T cell |
| MDS-12 | F | 24 | RCMD | NA | ND | 50MB | Bone marrow | CD3+ T cell |
| MDS-13 | F | 40 | del(5q) | 46,XX,del5q | ND | 38MB | Bone marrow | CD3+ T cell |
| MDS-14 | F | 46 | del(5q) | 46,XX,del5q | ND | 38MB | Bone marrow | CD3+ T cell |
| MDS-15 | M | 77 | RAEB-1 | 46,XY | 0.5 | 50MB | Bone marrow | CD3+ T cell |
| MDS-16 | M | 86 | RAEB-1 | 46,XY | 1 | 50MB | Bone marrow | CD3+ T cell |
| MDS-17 | M | 73 | RAEB-1 | 46,XY | 0.5 | 50MB | Bone marrow | CD3+ T cell |
| MDS-18 | F | 75 | RAEB-2 | 46,XX | 2 | 38MB | Bone marrow | CD3+ T cell |
| MDS-19 | F | 78 | RAEB-2 | 46,XX | 2 | 50MB | Bone marrow | CD3+ T cell |
| MDS-20 | F | 56 | RAEB-2 | 46,XX,t(6;9)(p23;q34) | ND | 38MB | Bone marrow | CD3+ T cell |
| MDS-21 | M | 59 | RAEB-2 | 46,XY | 2 | 50MB | Bone marrow | CD3+ T cell |
| tAML-01 | M | 77 | AML/MRC | 46,XY | NA | 38MB | Bone marrow | CD3+ T cell |
| tAML-02 | F | 68 | AML/MRC | 47,XX,der(1;7)(q10;p10),+8 | NA | 38MB | Peripheral blood(CD3-) | CD3+ T cell |
| tAML-03 | M | 67 | AML/MRC | Complex | NA | 38MB | Bone marrow | CD3+ T cell |
| tAML-04 | M | 58 | AML/MRC | NA | NA | 38MB | Bone marrow | CD3+ T cell |
| tAML-05 | M | 59 | AML/MRC | 47,XY,+11 | NA | 38MB | Bone marrow | CD3+ T cell |
| tAML-06 | M | 73 | AML/MRC | 45,XY,-7 | NA | 50MB | Bone marrow | CD3+ T cell |
| tAML-07 | F | 73 | AML/MRC | Complex | NA | 50MB | Bone marrow | Buccal mucosa |
| CMML-01 | M | 48 | CMML1 | 46,XY | ND | 38MB | Peripheral blood(CD3-) | CD3+ T cell |
| CMML-02 | M | 83 | CMML2 | 47,XY,+14 | ND | 38MB | Bone marrow | CD3+ T cell |
| CMML-03 | M | 77 | CMML1 | 46,XY | 0.5 | 50MB | Bone marrow | CD3+ T cell |
| CMML-04 | F | 88 | CMML1 | NA | ND | 50MB | Bone marrow | CD3+ T cell |

Abbreviations:RA, refractory anemia; RCMD, refractory cytopenia with multilineage dysplasia; del(5q), MDS with isolated del(5q);
RAEB-1, refractory anemia with excess blasts-1; RAEB-2, refractory anemia with excess blasts-2;
AML/MRC, acute myeloid leukemia with myelodysplasia related changes;
CMML1, chronic myelomonocytic leukemia 1; CMML2, chronic myelomonocytic leukemia 2;
NA, not available; ND, not determined.

**Supplementary Table 10. Primer sets used for qPCR analysis of NMD pathway genes**

| Gene name | Forward Primer | Reverse Primer |
|---|---|---|
| *UPF1* | 5'-CCCATTGCGAGGGGCCTTGG-3' | 5'-GTGTGCAGCGCTCCGTGTCT-3' |
| *UPF2* | 5'-GTTGGTACGGGCACTCTTCAT-3' | 5'-CCCCCTCAGCATGGAACAAA-3' |
| *UPF3B* | 5'-CTTCAGGGCAAAGAATAGAGAGA-3' | 5'-TTGACACAAGACTTACTCCTCTG-3' |
| *SMG1* | 5'-TCCAAACGTTGGGGAGGGGA-3' | 5'-TCCCATGGAAGGCAAACGACCC-3' |
| *SMG5* | 5'-CAAGGCAGCCAAAATGTACCAC-3' | 5'-GTAGGAGGCTTTGCAGATAC-3' |
| *SMG6* | 5'-TCCCTGAGGCCTGGACTCGC-3' | 5'-GTCCACGGGGACGCAACTGG-3' |
| *SMG7* | 5'-GCTGTCCCAGAGGGGAGGGG-3' | 5'-GCTGGGGTTGCCCAGACTGC-3' |
| *BTZ* | 5'-AACCTGGCCACTGTCCCTGT-3' | 5'-TGCCCACAGGCTTGAGGAGACA-3' |
| *DHX34* | 5'-TGCCCATCCACCTGGGTCACTT-3' | 5'-ATGCCCGGCTCGGCCTTTTA-3' |
| *GAPDH* | 5'-ACTGGCATGGCCTTCCGTGT-3' | 5'-ATGCCAGCCCCAGCGTCAAA-3' |

**Supplementary Table 2. The list of the mutated genes in whole-exome analysis**

| | Annotated Gene | Mutaion Type | RefSeq | Amino Acid Change | Chr | Position | Allele Change | Prediction by MutationTaster[#] | Sample Number |
|---|---|---|---|---|---|---|---|---|---|
| | *ADHFE1* | Missense | NM_144650 | A233G | Chr8 | 67361169 | C>G | Polymorphism | MDS-19 |
| | *AGAP1* | Missense | NM_001037131 | S636N | Chr2 | 236957718 | G>A | Disease_causing | tAML-05 |
| | *ALDH9A1* | Splice site | NM_000696 | Q310_splice | Chr1 | 165648678 | A>G | Disease_causing | MDS-20 |
| | *ANKRD34B* | Missense | NM_001004441 | H90Y | Chr5 | 79855571 | C>T | Disease_causing | MDS-19 |
| | *ANPEP* | Missense | NM_001150 | E355G | Chr15 | 90347599 | A>G | Disease_causing | tAML-03 |
| | *APBA2* | Missense | NM_005503 | V694I | Chr15 | 29406121 | G>A | Disease_causing | MDS-18 |
| | *APOL6* | Missense | NM_030641 | D150N | Chr22 | 36055059 | G>A | Polymorphism | tAML-02 |
| | *ARIH2* | Missense | NM_006321 | R484I | Chr3 | 49020672 | G>T | Disease_causing | MDS-11 |
| | *ARMC9* | Missense | NM_025139 | G48V | Chr2 | 232072931 | G>T | Polymorphism | MDS-07 |
| * | *ASXL1* | Nonsense | NM_015338 | R693X | Chr20 | 31022592 | C>T | Disease_causing | CMML-01 |
| * | *ASXL1* | Nonsense | NM_015338 | S852X | Chr20 | 31023070 | C>A | Disease_causing | MDS-19 |
| | *ATM* | Frameshift | NM_000051 | K25fs | Chr11 | 108098503 | */-A | Disease_causing | tAML-04 |
| | *ATP1A2* | Missense | NM_000702 | R564W | Chr1 | 160100250 | C>T | Disease_causing | tAML-04 |
| | *ATP6V0A2* | Missense | NM_012463 | K363R | Chr12 | 124228381 | A>G | Polymorphism | MDS-07 |
| * | *ATRX* | Missense | NM_138270 | V2124A | ChrX | 76814159 | T>C | Disease_causing | MDS-12 |
| | *BCOR* | Nonsense | NM_017745 | Q1187X | ChrX | 39923047 | C>T | Disease_causing | CMML-02 |
| | *BCOR* | Frameshift | NM_017745 | V1351fs | ChrX | 39922019 | */+C | Disease_causing | MDS-19 |
| | *BRCC3* | Nonsense | NM_001018055 | R81X | ChrX | 154305490 | C>T | Disease_causing | MDS-15 |
| | *BTBD10* | Missense | NM_032320 | H76Y | Chr11 | 13443261 | C>T | Disease_causing | MDS-21 |
| | *C10orf68* | Missense | NM_024688 | K541R | Chr10 | 33137644 | A>G | Polymorphism | MDS-20 |
| | *C11orf61* | Splice site | NM_024631 | D204_splice | Chr11 | 124642951 | G>C | Disease_causing | MDS-06 |
| | *C11orf9* | Frameshift | NM_001127392 | S261fs | Chr11 | 61539012 | */+C | Disease_causing | MDS-09 |
| | *C14orf145* | Missense | NM_152446 | R219Q | Chr14 | 81329207 | G>A | Polymorphism | tAML-04 |
| | *C14orf38* | Missense | NM_001164399 | T349A | Chr14 | 60012008 | A>G | Polymorphism | CMML-03 |
| | *C1orf101* | Splice site | NM_001130957 | G179_splice | Chr1 | 244682001 | G>A | Disease_causing | CMML-02 |
| | *C1QTNF3* | Missense | NM_030945 | P94L | Chr5 | 34033479 | C>T | Disease_causing | MDS-08 |
| | *C2orf67* | Missense | NM_152519 | P639L | Chr2 | 210908668 | C>T | Disease_causing | CMML-04 |
| | *C3orf63* | Missense | NM_015224 | E120V | Chr3 | 56681218 | A>T | Polymorphism | MDS-12 |
| | *C6orf170* | Nonsense | NM_152730 | G534X | Chr6 | 121602698 | G>T | Disease_causing | MDS-06 |
| | *CACNA1E* | Missense | NM_000721 | R372C | Chr1 | 181680148 | C>T | Disease_causing | tAML-07 |
| | *CACNB2* | Missense | NM_201572 | F215L | Chr10 | 18803967 | T>G | Polymorphism | tAML-04 |
| | *CALCR* | Missense | NM_001164737 | G418D | Chr7 | 93063605 | G>A | Disease_causing | MDS-09 |
| | *CASR* | Missense | NM_000388 | V621M | Chr3 | 122002662 | G>A | Disease_causing | MDS-08 |
| * | *CBL* | Missense | NM_005188 | W408L | Chr11 | 119149003 | G>T | Disease_causing | CMML-01 |
| | *CCDC83* | Missense | NM_173556 | E368K | Chr11 | 85627205 | G>A | Disease_causing | CMML-03 |
| | *CDH11* | Missense | NM_001797 | G247R | Chr16 | 65025743 | G>A | Disease_causing | MDS-07 |
| | *CHD7* | Missense | NM_017780 | M2509T | Chr8 | 61769365 | T>C | Polymorphism | MDS-15 |
| | *CHRM2* | Missense | NM_000739 | R52H | Chr7 | 136699767 | G>A | Disease_causing | MDS-14 |
| | *CKAP5* | Missense | NM_014756 | E615Q | Chr11 | 46811658 | G>C | Disease_causing | MDS-18 |
| | *CKLF* | Splice site | NM_016951 | - | Chr16 | 66592252 | */+T | Polymorphism | MDS-08 |
| | *CLEC1A* | Missense | NM_016511 | A111E | Chr12 | 10233895 | C>A | Polymorphism | MDS-18 |
| | *CNGA1* | Missense | NM_001142564 | R356K | Chr4 | 47939651 | G>A | Disease_causing | tAML-07 |
| | *COL24A1* | Missense | NM_152890 | G1014S | Chr1 | 86341004 | G>A | Disease_causing | MDS-15 |
| | *COL4A5* | Missense | NM_000495 | Q1582H | ChrX | 107938094 | G>C | Disease_causing | MDS-06 |
| | *CPAMD8* | Missense | NM_015692 | R558Q | Chr19 | 17091360 | G>A | Polymorphism | MDS-07 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *CRKL* | Missense | NM_005207 | G147V | Chr22 | 21288195 | G>T | Disease_causing | MDS-08 |
| | *CSMD1* | Missense | NM_033225 | G2681S | Chr8 | 2857642 | G>A | Disease_causing | MDS-19 |
| | *CSMD2* | Missense | NM_052896 | V3201I | Chr1 | 33998788 | G>A | Polymorphism | CMML-01 |
| | *CYP7B1* | Nonsense | NM_004820 | R63X | Chr8 | 65537032 | C>T | Disease_causing | CMML-03 |
| | *DACH1* | Missense | NM_080760 | E248Q | Chr13 | 72440166 | G>C | Disease_causing | MDS-12 |
| | *DAPK1* | Missense | NM_004938 | F840L | Chr9 | 90312026 | T>C | Disease_causing | MDS-03 |
| | *DDX4* | Nonsense | NM_001166534 | W521X | Chr5 | 55111164 | G>A | Disease_causing | MDS-04 |
| | *DENND2A* | Missense | NM_015689 | P312R | Chr7 | 140301263 | C>G | Disease_causing | tAML-02 |
| * | *DNMT3A* | Missense | NM_175629 | R882H | Chr2 | 25457242 | G>A | Disease_causing | CMML-02 |
| * | *DNMT3A* | Missense | NM_175629 | G543C | Chr2 | 25467449 | G>T | Disease_causing | MDS-19 |
| | *DOCK4* | Missense | NM_014705 | V770M | Chr7 | 111503593 | G>A | Disease_causing | tAML-07 |
| | *DOT1L* | Nonsense | NM_032482 | Q430X | Chr19 | 2210791 | C>T | Disease_causing | MDS-08 |
| | *DRD1* | Missense | NM_000794 | Y131H | Chr5 | 174869712 | T>C | Disease_causing | MDS-12 |
| | *DYNC1H1* | Missense | NM_001376 | G1432C | Chr14 | 102467590 | G>T | Disease_causing | MDS-12 |
| | *EDC4* | Missense | NM_014329 | R473W | Chr16 | 67912989 | C>T | Disease_causing | MDS-20 |
| | *EGF* | Missense | NM_001178131 | C887Y | Chr4 | 110914454 | G>A | Disease_causing | MDS-12 |
| | *EIF2AK2* | Missense | NM_001135651 | N302K | Chr2 | 37353434 | C>G | Disease_causing | MDS-10 |
| | *EIF4G3* | Missense | NM_003760 | R988K | Chr1 | 21186991 | G>A | Disease_causing | CMML-03 |
| | *ELANE* | Missense | NM_001972 | A216T | Chr19 | 856006 | G>A | Polymorphism | tAML-01 |
| | *EML1* | Missense | NM_004434 | E184G | Chr14 | 100360969 | A>G | Disease_causing | MDS-19 |
| | *EML4* | Missense | NM_019063 | I356V | Chr2 | 42513463 | A>G | Polymorphism | tAML-01 |
| | *ENC1* | Splice site | NM_003633 | M1_spilice | Chr5 | 73932323 | G>A | Disease_causing | MDS-12 |
| | *EPAS1* | Missense | NM_001430 | W639L | Chr2 | 46607727 | G>T | Disease_causing | CMML-03 |
| | *ERBB4* | Missense | NM_005235 | R1067Q | Chr2 | 212251859 | G>A | Polymorphism | CMML-02 |
| | *ETNK1* | Missense | NM_018638 | N244S | Chr12 | 22811995 | A>G | Disease_causing | MDS-08 |
| | *ETV6* | Missense | NM_001987 | R339I | Chr12 | 12037385 | G>T | Disease_causing | tAML-02 |
| * | *EZH2* | Frameshift | NM_004456 | K718fs | Chr7 | 148506205 | */-AA | Disease_causing | CMML-01 |
| * | *EZH2* | Missense | NM_004456 | D664A | Chr7 | 148507463 | A>C | Disease_causing | tAML-01 |
| | *F5* | Missense | NM_000130 | K331E | Chr1 | 169524547 | A>G | Polymorphism | MDS-07 |
| | *FAM13A* | Missense | NM_014883 | S464C | Chr4 | 89702384 | A>T | Polymorphism | tAML-02 |
| | *FAM175A* | Missense | NM_139076 | S126L | Chr4 | 84391455 | C>T | Disease_causing | MDS-11 |
| | *FAM47C* | Missense | NM_001013736 | H235R | ChrX | 37027187 | A>G | Polymorphism | MDS-09 |
| | *FAM83B* | Nonsense | NM_001010872 | G513X | Chr6 | 54805306 | G>T | Disease_causing | tAML-02 |
| | *FAT2* | Frameshift | NM_001447 | P2583fs | Chr5 | 150922940 | */-C | Disease_causing | MDS-12 |
| | *FBXL20* | Missense | NM_001184906 | N331K | Chr17 | 37420542 | C>G | Disease_causing | MDS-11 |
| | *FBXO11* | Missense | NM_001190274 | W621C | Chr2 | 48046152 | G>T | Disease_causing | tAML-01 |
| | *FER1L6* | Missense | NM_001039112 | T1195I | Chr8 | 125078697 | C>T | Disease_causing | MDS-11 |
| | *FILIP1* | Missense | NM_015687 | A319V | Chr6 | 76024592 | C>T | Disease_causing | tAML-01 |
| * | *FLT3* | Missense | NM_004119 | Q575P | Chr13 | 28608332 | A>C | Disease_causing | tAML-01 |
| | *FMN2* | Missense | NM_020066 | I1714T | Chr1 | 240635752 | T>C | Disease_causing | CMML-01 |
| | *FSIP2* | Missense | NM_173651 | M1246I | Chr2 | 186655334 | G>A | Polymorphism | MDS-12 |
| | *FYB* | Missense | NM_001465 | P260A | Chr5 | 39202285 | C>G | Polymorphism | MDS-16 |
| | *GLE1* | Missense | NM_001499 | R295Q | Chr9 | 131286112 | G>A | Disease_causing | tAML-01 |
| | *GNB1* | Missense | NM_002074 | K57E | Chr1 | 1747229 | A>G | Disease_causing | tAML-03 |
| | *GNL2* | Missense | NM_013285 | T642A | Chr1 | 38033903 | A>G | Polymorphism | CMML-01 |
| | *GRID1* | Nonsense | NM_017551 | W501X | Chr10 | 87487642 | G>A | Disease_causing | MDS-09 |
| | *GRIN2B* | Missense | NM_000834 | V1189M | Chr12 | 13716607 | G>A | Polymorphism | tAML-07 |
| | *GTF3C4* | Frameshift | NM_012204 | M614fs | Chr9 | 135554847 | */+G | Disease_causing | MDS-19 |
| | *GUCA1A* | Missense | NM_000409 | L98F | Chr6 | 42146108 | C>T | Disease_causing | tAML-07 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *HDAC9* | Missense | NM_178425 | V650I | Chr7 | 18788666 | G>A | Polymorphism | CMML-03 |
| | *HELLS* | Frameshift | NM_018063 | P508fs | Chr10 | 96350205 | */-A | Disease_causing | tAML-02 |
| | *HIVEP1* | Missense | NM_002114 | V489I | Chr6 | 12121493 | G>A | Polymorphism | MDS-18 |
| | *HIVEP2* | Missense | NM_006734 | R1337Q | Chr6 | 143091866 | G>A | Disease_causing | MDS-06 |
| | *HSPH1* | Missense | NM_006644 | P453S | Chr13 | 31722117 | C>T | Disease_causing | MDS-17 |
| | *HTR1A* | Missense | NM_000524 | N300T | Chr5 | 63256648 | A>C | Polymorphism | tAML-07 |
| | *HTR2A* | Missense | NM_001165947 | C143F | Chr13 | 47409708 | G>T | Disease_causing | CMML-04 |
| | *HYDIN* | Missense | NM_032821 | G5034R | Chr16 | 70841746 | G>A | Disease_causing | CMML-02 |
| * | *IDH1* | Missense | NM_005896 | R132C | Chr2 | 209113113 | C>T | Disease_causing | tAML-07 |
| * | *IDH2* | Missense | NM_002168 | R140Q | Chr15 | 90631934 | G>A | Disease_causing | MDS-18 |
| | *IGSF3* | Missense | NM_001007237 | D356H | Chr1 | 117150720 | G>C | Disease_causing | tAML-07 |
| | *IMP4* | Missense | NM_033416 | R76L | Chr2 | 131102979 | G>T | Disease_causing | tAML-02 |
| | *IRS4* | Missense | NM_003604 | S513N | ChrX | 107978037 | G>A | Polymorphism | MDS-19 |
| | *ITGAD* | Missense | NM_005353 | L798P | Chr16 | 31427861 | T>C | Polymorphism | tAML-03 |
| | *ITPKC* | Missense | NM_025194 | G520E | Chr19 | 41239154 | G>A | Disease_causing | CMML-01 |
| | *KANK4* | Missense | NM_181712 | S917G | Chr1 | 62713278 | A>G | Polymorphism | CMML-01 |
| | *KBTBD4* | Frameshift | NM_016506 | E386fs | Chr11 | 47594882 | */+G | Disease_causing | MDS-12 |
| | *KBTBD5* | Missense | NM_152393 | P539L | Chr3 | 42732359 | C>T | Polymorphism | MDS-07 |
| | *KCNQ3* | Missense | NM_004519 | A669T | Chr8 | 133142123 | G>A | Polymorphism | MDS-06 |
| | *KCTD1* | Missense | NM_001142730 | S605I | Chr18 | 24081210 | G>T | Polymorphism | tAML-04 |
| | *KIAA0232* | Missense | NM_014743 | V622M | Chr4 | 6863973 | G>A | Polymorphism | MDS-15 |
| | *KIAA1217* | Missense | NM_001098501 | S588R | Chr10 | 24790342 | T>G | Polymorphism | tAML-02 |
| | *KIAA1522* | Missense | NM_001198973 | S138P | Chr1 | 33238468 | T>C | Polymorphism | MDS-19 |
| * | *KRAS* | Missense | NM_033360 | K5E | Chr12 | 25398306 | A>G | Disease_causing | CMML-01 |
| * | *KRAS* | Missense | NM_033360 | G12D | Chr12 | 25398284 | G>A | Disease_causing | tAML-06 |
| | *KRT14* | Splice site | NM_000526 | K309_splice | Chr17 | 39740011 | G>A | Disease_causing | tAML-02 |
| | *KRT37* | Missense | NM_003770 | R377L | Chr17 | 39577730 | G>T | Polymorphism | MDS-07 |
| | *KRT39* | Missense | NM_213656 | Q204K | Chr17 | 39119977 | C>A | Polymorphism | CMML-04 |
| | *KRT72* | Missense | NM_001146226 | F259L | Chr12 | 52986201 | C>A | Polymorphism | MDS-12 |
| | *L1CAM* | Missense | NM_000425 | V1215M | ChrX | 153128249 | G>A | Polymorphism | tAML-03 |
| | *LAMA1* | Missense | NM_005559 | A558T | Chr18 | 7037642 | G>A | Polymorphism | MDS-04 |
| | *LAMA5* | Missense | NM_005560 | R2835K | Chr20 | 60889360 | G>A | Disease_causing | tAML-01 |
| | *LAMB4* | Nonsense | NM_007356 | W932X | Chr7 | 107706248 | G>A | Disease_causing | MDS-19 |
| | *LARP1B* | Missense | NM_018078 | Q524R | Chr4 | 129083395 | A>G | Polymorphism | MDS-11 |
| | *LCA5* | Missense | NM_001122769 | Q123H | Chr6 | 80223280 | G>T | Polymorphism | CMML-04 |
| | *LIX1* | Missense | NM_153234 | I211N | Chr5 | 96430669 | T>A | Disease_causing | tAML-07 |
| | *LPHN3* | Missense | NM_015236 | F951L | Chr4 | 62849142 | C>A | Disease_causing | MDS-12 |
| | *LRP1* | Missense | NM_002332 | D152V | Chr12 | 57538761 | A>T | Disease_causing | MDS-03 |
| | *LRRC37A3* | Missense | NM_199340 | A1025T | Chr17 | 62859117 | G>A | Polymorphism | CMML-02 |
| | *LRRC46* | Missense | NM_033413 | D235N | Chr17 | 45914223 | G>A | Polymorphism | MDS-04 |
| | *LRRFIP1* | Missense | NM_001137552 | R79C | Chr2 | 238629451 | C>T | Disease_causing | MDS-06 |
| | *LRSAM1* | Missense | NM_001005374 | D32G | Chr9 | 130217299 | A>G | Disease_causing | MDS-08 |
| | *LTBP2* | Frameshift | NM_000428 | H1816fs | Chr14 | 74967606 | */+C | Disease_causing | MDS-16 |
| | *LUC7L2* | Splice site | NM_016019 | - | Chr7 | 139060903 | */+T | Disease_causing | CMML-01 |
| | *LY75* | Missense | NM_001198760 | G1013S | Chr2 | 160706604 | G>A | Disease_causing | CMML-02 |
| | *MAGEB1* | Missense | NM_002363 | L150F | ChrX | 30269058 | C>T | Polymorphism | tAML-01 |
| | *MANBA* | Missense | NM_005908 | I333V | Chr4 | 103595191 | A>G | Disease_causing | MDS-09 |
| | *MAST4* | Missense | NM_001164664 | K1799R | Chr5 | 66460403 | A>G | Polymorphism | tAML-07 |
| | *MC5R* | Missense | NM_005913 | V202G | Chr18 | 13826369 | T>G | Disease_causing | MDS-04 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *MCRS1* | Frameshift | NM_001012300 | D289fs | Chr12 | 49954132 | */+G | Disease_causing | MDS-07 |
| | *MED12* | Frameshift | NM_005120 | S1933fs | ChrX | 70357456 | */-A | Disease_causing | tAML-07 |
| | *MID2* | Missense | NM_012216 | C207R | ChrX | 107084514 | T>C | Disease_causing | CMML-01 |
| | *MKRN3* | Missense | NM_005664 | R469Q | Chr15 | 23812335 | G>A | Polymorphism | tAML-04 |
| | *MPP2* | Missense | NM_005374 | A452V | Chr17 | 41956770 | C>T | Disease_causing | MDS-16 |
| | *MSLNL* | Missense | NM_001025190 | S1024L | Chr16 | 819519 | C>T | Polymorphism | MDS-20 |
| | *MTUS2* | Missense | NM_001033602 | R1362H | Chr13 | 30077288 | G>A | Disease_causing | tAML-05 |
| | *MYO1H* | Missense | NM_001101421 | E1002V | Chr12 | 109883377 | A>T | NA | CMML-03 |
| | *MYST1* | Missense | NM_032188 | Y90C | Chr16 | 31131564 | A>G | Disease_causing | CMML-02 |
| | *NCAPH* | Splice site | NM_015341 | K151_splice | Chr2 | 97008587 | G>A | Disease_causing | tAML-02 |
| | *NONO* | Frameshift | NM_001145409 | Y267fs | ChrX | 70516754 | */+T | NA | MDS-07 |
| * | *NRAS* | Missense | NM_002524 | G13D | Chr1 | 115258744 | G>A | Disease_causing | MDS-19 |
| * | *NRAS* | Missense | NM_002524 | G12S | Chr1 | 115258748 | G>A | Disease_causing | MDS-19 |
| * | *NRAS* | Missense | NM_002524 | G13D | Chr1 | 115258744 | G>A | Disease_causing | MDS-20 |
| * | *NRAS* | Missense | NM_002524 | G12S | Chr1 | 115258748 | G>A | Disease_causing | tAML-05 |
| | *NRIP1* | Frameshift | NM_003489 | T603fs | Chr21 | 16338706 | */-C | Disease_causing | MDS-20 |
| | *NRP1* | Missense | NM_003873 | A590P | Chr10 | 33491915 | G>C | Disease_causing | MDS-18 |
| | *NXF1* | Missense | NM_006362 | D438E | Chr11 | 62563790 | T>G | Disease_causing | MDS-18 |
| | *OR1I1* | Missense | NM_001004713 | G242D | Chr19 | 15198601 | G>A | Polymorphism | tAML-05 |
| | *OR51S1* | Missense | NM_001004758 | N312S | Chr11 | 4869504 | A>G | Polymorphism | tAML-07 |
| | *OR52H1* | Missense | NM_001005289 | V219L | Chr11 | 5566099 | G>T | Polymorphism | tAML-04 |
| | *OR8H1* | Missense | NM_001005199 | G95D | Chr11 | 56058255 | G>A | Polymorphism | tAML-01 |
| | *PAMR1* | Missense | NM_015430 | R550Q | Chr11 | 35456088 | G>A | Disease_causing | tAML-06 |
| | *PANX3* | Missense | NM_052959 | R296Q | Chr11 | 124489539 | G>A | Polymorphism | CMML-02 |
| | *PAPPA2* | Missense | NM_021936 | S718N | Chr1 | 176659288 | G>A | Polymorphism | MDS-19 |
| | *PCDHAC1* | Missense | NM_031882 | R622C | Chr5 | 140308341 | C>T | Polymorphism | MDS-15 |
| | *PCDHB7* | Missense | NM_018940 | R229C | Chr5 | 140553101 | C>T | Polymorphism | tAML-02 |
| | *PCDHGC3* | Frameshift | NM_032402 | T659fs | Chr5 | 140857660 | */-T | Disease_causing | CMML-01 |
| | *PDS5B* | Frameshift | NM_015032 | G721fs | Chr13 | 33306277 | */+C | Disease_causing | tAML-02 |
| | *PDZRN3* | Missense | NM_015009 | S891L | Chr3 | 73433045 | C>T | Disease_causing | MDS-07 |
| | *PHF6* | Nonsense | NM_032458 | G291X | ChrX | 133551235 | G>T | Disease_causing | MDS-18 |
| | *PHF8* | Splice site | NM_015107 | I814_splice | ChrX | 53989373 | G>A | Disease_causing | tAML-02 |
| | *PLCXD3* | Missense | NM_001005473 | P213L | Chr5 | 41382102 | C>T | Disease_causing | MDS-12 |
| | *POLR2A* | Missense | NM_000937 | S1138L | Chr17 | 7411742 | C>T | Disease_causing | tAML-02 |
| | *PRAMEF11* | Missense | NM_001146344 | R104W | Chr1 | 12887547 | C>T | Polymorphism | MDS-11 |
| | *PRPF40B* | Missense | NM_001031698 | P383L | Chr12 | 50029195 | C>T | Disease_causing | tAML-04 |
| | *PTX4* | Nonsense | NM_001013658 | R290X | Chr16 | 1536494 | C>T | Disease_causing | tAML-01 |
| | *PVRL4* | Missense | NM_030916 | V346M | Chr1 | 161044128 | G>A | Polymorphism | MDS-19 |
| | *RABEP1* | Missense | NM_004703 | R685H | Chr17 | 5280439 | G>A | Disease_causing | MDS-16 |
| | *RAI2* | Missense | NM_021785 | A396P | ChrX | 17818945 | G>C | Polymorphism | tAML-05 |
| | *RFC1* | Nonsense | NM_002913 | E234X | Chr4 | 39324980 | G>T | Disease_causing | MDS-11 |
| | *RFX6* | Missense | NM_173560 | A561V | Chr6 | 117246619 | C>T | Disease_causing | MDS-07 |
| | *ROCK2* | Frameshift | NM_004850 | Q844fs | Chr2 | 11347871 | */+A | Disease_causing | MDS-12 |
| * | *RUNX1* | Frameshift | NM_001001890 | L294fs | Chr21 | 36171602 | */+A | Disease_causing | MDS-19 |
| * | *RUNX1* | Missense | NM_001001890 | D171N | Chr21 | 36231792 | G>A | Disease_causing | tAML-02 |
| | *SAA4* | Missense | NM_006512 | N94K | Chr11 | 18253160 | C>G | Polymorphism | tAML-02 |
| | *SCML2* | Missense | NM_006089 | F452I | ChrX | 18275070 | T>A | Polymorphism | tAML-02 |
| | *SEC14L4* | Missense | NM_174977 | Q406H | Chr22 | 30886097 | G>C | Polymorphism | MDS-20 |
| | *SEZ6L* | Missense | NM_001184773 | T710M | Chr22 | 26736515 | C>T | Disease_causing | tAML-07 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *SF3A1* | Missense | NM_001005409 | K166T | Chr22 | 30738828 | A>C | Disease_causing | MDS-07 |
| | *SF3B1* | Missense | NM_012433 | K700E | Chr2 | 198266834 | A>G | Disease_causing | MDS-21 |
| | *SRSF2* | Missense | NM_003016 | P95H | Chr17 | 74732959 | C>A | Disease_causing | CMML-04 |
| | *SRSF2* | Missense | NM_003016 | P95L | Chr17 | 74732959 | C>T | Disease_causing | MDS-09 |
| | *SRSF2* | Missense | NM_003016 | P95H | Chr17 | 74732959 | C>A | Disease_causing | MDS-18 |
| | *SH3TC2* | Missense | NM_024577 | A1090V | Chr5 | 148389891 | C>T | Disease_causing | tAML-01 |
| | *SKIV2L2* | Missense | NM_015360 | P720T | Chr5 | 54683907 | C>A | Disease_causing | CMML-04 |
| | *SLC20A1* | Missense | NM_005415 | P411H | Chr2 | 113416964 | C>A | Disease_causing | MDS-11 |
| | *SLC9A10* | Missense | NM_183061 | R34G | Chr3 | 111999619 | C>G | Polymorphism | MDS-15 |
| | *SLCO4C1* | Missense | NM_180991 | S190L | Chr5 | 101627097 | C>T | Polymorphism | MDS-09 |
| | *SLIT3* | Missense | NM_003062 | G1440S | Chr5 | 168096806 | G>A | Disease_causing | CMML-02 |
| | *SORBS2* | Frameshift | NM_001145675 | E436fs | Chr4 | 186515043 | */+G | Disease_causing | MDS-15 |
| | *SPATA20* | Missense | NM_022827 | R698H | Chr17 | 48631747 | G>A | Polymorphism | tAML-05 |
| | *SPATS1* | Missense | NM_145026 | H109R | Chr6 | 44328221 | A>G | Polymorphism | tAML-06 |
| | *SPRYD5* | Nonsense | NM_032681 | Y21X | Chr11 | 55652967 | C>A | Disease_causing | CMML-03 |
| | *SRCAP* | Missense | NM_006662 | D298N | Chr16 | 30721207 | G>A | Polymorphism | MDS-16 |
| | *SRL* | Missense | NM_001098814 | G193D | Chr16 | 4245586 | G>A | Disease_causing | MDS-12 |
| | *SSX7* | Missense | NM_173358 | R169H | ChrX | 52674554 | G>A | Polymorphism | tAML-07 |
| | *STAC3* | Frameshift | NM_145064 | K234fs | Chr12 | 57638717 | */-G | Disease_causing | MDS-11 |
| | *STAG1* | Missense | NM_005862 | H1085Y | Chr3 | 136068018 | C>T | Disease_causing | MDS-11 |
| | *STAG2* | Nonsense | NM_001042749 | R1012X | ChrX | 123217380 | C>T | Disease_causing | MDS-19 |
| | *STEAP2* | Missense | NM_152999 | G53S | Chr7 | 89854553 | G>A | Disease_causing | CMML-04 |
| | *SYT10* | Missense | NM_198992 | M393I | Chr12 | 33538125 | G>A | Disease_causing | CMML-04 |
| | *SYTL2* | Missense | NM_032943 | A611V | Chr11 | 85420470 | C>T | Disease_causing | CMML-04 |
| | *TBC1D9B* | Nonsense | NM_015043 | R555X | Chr5 | 179305428 | C>T | Disease_causing | MDS-18 |
| | *TBX19* | Missense | NM_005149 | D47G | Chr1 | 168250468 | A>G | Disease_causing | MDS-03 |
| | *TBX5* | Missense | NM_080717 | P384L | Chr12 | 114793593 | C>T | Polymorphism | tAML-05 |
| | *TCN1* | Missense | NM_001062 | N343S | Chr11 | 59622218 | A>G | Polymorphism | CMML-04 |
| | *TEP1* | Missense | NM_007110 | L1997P | Chr14 | 20845537 | T>C | Polymorphism | MDS-08 |
| * | *TET2* | Frameshift | NM_017628 | Q749fs | Chr4 | 106157345 | */-AAAT | Disease_causing | CMML-01 |
| * | *TET2* | Nonsense | NM_017628 | Q317X | Chr4 | 106156048 | C>T | Disease_causing | CMML-03 |
| * | *TET2* | Frameshift | NM_017628 | I340fs | Chr4 | 106156119 | */-C | Disease_causing | MDS-06 |
| * | *TET2* | Missense | NM_017628 | C1271W | Chr4 | 106180785 | C>G | Disease_causing | MDS-15 |
| * | *TET2* | Nonsense | NM_017628 | Y1255fs | Chr4 | 106164895 | */+A | Disease_causing | MDS-16 |
| * | *TET2* | Nonsense | NM_017628 | Q323X | Chr4 | 106156066 | C>T | Disease_causing | MDS-16 |
| * | *TET2* | Frameshift | NM_017628 | L1046fs | Chr4 | 106158237 | */+A | Disease_causing | MDS-19 |
| * | *TET2* | Nonsense | NM_017628 | S825X | Chr4 | 106157573 | C>G | Disease_causing | MDS-19 |
| * | *TET2* | Frameshift | NM_017628 | H994fs | Chr4 | 106158081 | */-C | Disease_causing | tAML-01 |
| * | *TET2* | Nonsense | NM_017628 | R550X | Chr4 | 106156747 | C>T | Disease_causing | tAML-01 |
| * | *TET2* | Frameshift | NM_017628 | S402fs | Chr4 | 106156304 | */+A | Disease_causing | tAML-07 |
| * | *TET2* | Nonsense | NM_017628 | Q1680X | Chr4 | 106196705 | C>T | Disease_causing | tAML-07 |
| | *TMEM146* | Missense | NM_152784 | W663L | Chr19 | 5776218 | G>T | Polymorphism | MDS-18 |
| | *TMEM87B* | Missense | NM_032824 | M448I | Chr2 | 112856243 | G>A | Disease_causing | CMML-04 |
| | *TNK2* | Missense | NM_001010938 | N94S | Chr3 | 195615368 | A>G | Disease_causing | MDS-08 |
| | *TNKS1BP1* | Missense | NM_033396 | K1083N | Chr11 | 57076936 | G>C | Polymorphism | MDS-07 |
| | *TNS4* | Missense | NM_032865 | A547V | Chr17 | 38638413 | C>T | Disease_causing | tAML-07 |
| * | *TP53* | Missense | NM_001126116 | R116W | Chr17 | 7577539 | C>T | Disease_causing | MDS-09 |
| * | *TP53* | Frameshift | NM_001126116 | M1fs | Chr17 | 7578532 | */-T | Disease_causing | tAML-03 |
| | *TP53BP1* | Nonsense | NM_001141980 | L807X | Chr15 | 43748386 | T>G | Disease_causing | MDS-07 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *TRPC4* | Missense | NM_001135957 | R491H | Chr13 | 38237769 | G>A | Disease_causing | CMML-02 |
| *TRPM4* | Missense | NM_001195227 | V801M | Chr19 | 49703925 | G>A | Polymorphism | tAML-06 |
| *TTF2* | Missense | NM_003594 | R498H | Chr1 | 117619365 | G>A | Polymorphism | MDS-03 |
| *TTN* | Missense | NM_133432 | G19774R | Chr2 | 179424719 | G>A | Disease_causing | MDS-11 |
| *U2AF35* | Missense | NM_006758 | Q157R | Chr21 | 44514777 | A>G | Disease_causing | CMML-01 |
| *U2AF35* | Missense | NM_006758 | S34F | Chr21 | 44524456 | C>T | Disease_causing | CMML-02 |
| *U2AF35* | Missense | NM_006758 | Q157R | Chr21 | 44514777 | A>G | Disease_causing | MDS-03 |
| *U2AF35* | Missense | NM_006758 | A26V | Chr21 | 44524480 | C>T | Disease_causing | MDS-12 |
| *U2AF35* | Missense | NM_006758 | Q157P | Chr21 | 44514777 | A>C | Disease_causing | MDS-15 |
| *UBE2Q1* | Frameshift | NM_017582 | N412fs | Chr1 | 154523415 | */+A | Disease_causing | MDS-12 |
| *UGCG* | Missense | NM_003358 | I363V | Chr9 | 114695179 | A>G | Disease_causing | CMML-04 |
| *UNC5D* | Missense | NM_080872 | R783W | Chr8 | 35624453 | C>T | Polymorphism | CMML-02 |
| *VPS13B* | Missense | NM_017890 | I935L | Chr8 | 100287461 | A>C | Polymorphism | tAML-05 |
| *VPS13D* | Missense | NM_015378 | A2597D | Chr1 | 12382678 | C>A | Disease_causing | CMML-03 |
| *ZKSCAN1* | Missense | NM_003439 | E118K | Chr7 | 99621481 | G>A | Disease_causing | MDS-09 |
| *ZMYND8* | Missense | NM_012408 | R35T | Chr20 | 45938930 | G>C | Polymorphism | tAML-02 |
| *ZNF211* | Missense | NM_198855 | G347V | Chr19 | 58152894 | G>T | Polymorphism | tAML-07 |
| *ZNF225* | Missense | NM_013362 | C349Y | Chr19 | 44635813 | G>A | NA | MDS-19 |
| *ZNF254* | Missense | NM_203282 | G216V | Chr19 | 24309449 | G>T | Polymorphism | tAML-07 |
| *ZNF518A* | Missense | NM_014803 | D490E | Chr10 | 97917549 | C>A | Polymorphism | MDS-12 |
| *ZNF534* | Missense | NM_001143939 | T418S | Chr19 | 52941927 | C>G | Polymorphism | MDS-12 |
| *ZNF583* | Missense | NM_001159861 | V101A | Chr19 | 56934329 | T>C | Polymorphism | MDS-20 |
| *ZRSR2* | Missense | NM_005089 | F239V | ChrX | 15833957 | T>G | Disease_causing | CMML-03 |
| *ZRSR2* | Splice site | NM_005089 | K257_splice | ChrX | 15834014 | G>A | Disease_causing | MDS-06 |
| *ZRSR2* | Frameshift | NM_005089 | G323fs | ChrX | 15840883 | */-G | Disease_causing | MDS-08 |
| *ZRSR2* | Nonsense | NM_005089 | E362X | ChrX | 15841000 | G>T | Disease_causing | MDS-16 |
| *ZRSR2* | Missense | NM_005089 | N261Y | ChrX | 15836719 | A>T | Disease_causing | tAML-05 |

** indicates that mutations of these genes have been reported previously in MDS.*

*# Pathogeneity of somatic mutations was predicted using Mutation Taster (http://www.mutationtaster.org/)*

Abbreviations: NA, not availble.

**Supplementary Table 3. Nonsense and frameshift mutations detected in whole-exome analysis**

| Annotated Gene | Mutaion Type | RefSeq | Amino Acid Change | Chr | Position | Allele Change | Sample Number |
|---|---|---|---|---|---|---|---|
| *ASXL1* | Nonsense | NM_015338 | R693X | Chr20 | 31022592 | C>T | CMML-01 |
| *ASXL1* | Nonsense | NM_015338 | S852X | Chr20 | 31023070 | C>A | MDS-19 |
| *ATM* | Frameshift | NM_000051 | K25fs | Chr11 | 108098503 | */-A | tAML-04 |
| *BCOR* | Nonsense | NM_017745 | Q1187X | ChrX | 39923047 | C>T | CMML-02 |
| *BCOR* | Frameshift | NM_017745 | V1351fs | ChrX | 39922019 | */+C | MDS-19 |
| *BRCC3* | Nonsense | NM_001018055 | R81X | ChrX | 154305490 | C>T | MDS-15 |
| *C11orf9* | Frameshift | NM_001127392 | S261fs | Chr11 | 61539012 | */+C | MDS-09 |
| *C6orf170* | Nonsense | NM_152730 | G534X | Chr6 | 121602698 | G>T | MDS-06 |
| *CYP7B1* | Nonsense | NM_004820 | R63X | Chr8 | 65537032 | C>T | CMML-03 |
| *DDX4* | Nonsense | NM_001166534 | W521X | Chr5 | 55111164 | G>A | MDS-04 |
| *DOT1L* | Nonsense | NM_032482 | Q430X | Chr19 | 2210791 | C>T | MDS-08 |
| *EZH2* | Frameshift | NM_004456 | K718fs | Chr7 | 148506205 | */-AA | CMML-01 |
| *FAM83B* | Nonsense | NM_001010872 | G513X | Chr6 | 54805306 | G>T | tAML-02 |
| *FAT2* | Frameshift | NM_001447 | P2583fs | Chr5 | 150922940 | */-C | MDS-12 |
| *GRID1* | Nonsense | NM_017551 | W501X | Chr10 | 87487642 | G>A | MDS-09 |
| *GTF3C4* | Frameshift | NM_012204 | M614fs | Chr9 | 135554847 | */+G | MDS-19 |
| *HELLS* | Frameshift | NM_018063 | P508fs | Chr10 | 96350205 | */-A | tAML-02 |
| *KBTBD4* | Frameshift | NM_016506 | E386fs | Chr11 | 47594882 | */+G | MDS-12 |
| *LAMB4* | Nonsense | NM_007356 | W932X | Chr7 | 107706248 | G>A | MDS-19 |
| *LTBP2* | Frameshift | NM_000428 | H1816fs | Chr14 | 74967606 | */+C | MDS-16 |
| *MCRS1* | Frameshift | NM_001012300 | D289fs | Chr12 | 49954132 | */+G | MDS-07 |
| *MED12* | Frameshift | NM_005120 | S1933fs | ChrX | 70357456 | */-A | tAML-07 |
| *NONO* | Frameshift | NM_001145409 | Y267fs | ChrX | 70516754 | */+T | MDS-07 |
| *NRIP1* | Frameshift | NM_003489 | T603fs | Chr21 | 16338706 | */-C | MDS-20 |
| *PCDHGC3* | Frameshift | NM_032402 | T659fs | Chr5 | 140857660 | */-T | CMML-01 |
| *PDS5B* | Frameshift | NM_015032 | G721fs | Chr13 | 33306277 | */+C | tAML-02 |
| *PHF6* | Nonsense | NM_032458 | G291X | ChrX | 133551235 | G>T | MDS-18 |
| *PTX4* | Nonsense | NM_001013658 | R290X | Chr16 | 1536494 | C>T | tAML-01 |
| *RFC1* | Nonsense | NM_002913 | E234X | Chr4 | 39324980 | G>T | MDS-11 |
| *ROCK2* | Frameshift | NM_004850 | Q844fs | Chr2 | 11347871 | */+A | MDS-12 |
| *RUNX1* | Frameshift | NM_001001890 | L294fs | Chr21 | 36171602 | */+A | MDS-19 |
| *SORBS2* | Frameshift | NM_001145675 | E436fs | Chr4 | 186515043 | */+G | MDS-15 |
| *SPRYD5* | Nonsense | NM_032681 | Y21X | Chr11 | 55652967 | C>A | CMML-03 |
| *STAC3* | Frameshift | NM_145064 | K234fs | Chr12 | 57638717 | */-G | MDS-11 |
| *STAG2* | Nonsense | NM_001042749 | R1012X | ChrX | 123217380 | C>T | MDS-19 |
| *TBC1D9B* | Nonsense | NM_015043 | R555X | Chr5 | 179305428 | C>T | MDS-18 |
| *TET2* | Frameshift | NM_017628 | Q749fs | Chr4 | 106157345 | */-AAAT | CMML-01 |
| *TET2* | Nonsense | NM_017628 | Q317X | Chr4 | 106156048 | C>T | CMML-03 |
| *TET2* | Frameshift | NM_017628 | I340fs | Chr4 | 106156119 | */-C | MDS-06 |
| *TET2* | Nonsense | NM_017628 | Q323X | Chr4 | 106156066 | C>T | MDS-16 |
| *TET2* | Nonsense | NM_017628 | Y1255fs | Chr4 | 106164895 | */+A | MDS-16 |
| *TET2* | Nonsense | NM_017628 | S825X | Chr4 | 106157573 | C>G | MDS-19 |
| *TET2* | Frameshift | NM_017628 | L1046fs | Chr4 | 106158237 | */+A | MDS-19 |
| *TET2* | Nonsense | NM_017628 | R550X | Chr4 | 106156747 | C>T | tAML-01 |
| *TET2* | Frameshift | NM_017628 | H994fs | Chr4 | 106158081 | */-C | tAML-01 |
| *TET2* | Frameshift | NM_017628 | S402fs | Chr4 | 106156304 | */+A | tAML-07 |
| *TET2* | Nonsense | NM_017628 | Q1680X | Chr4 | 106196705 | C>T | tAML-07 |
| *TP53* | Frameshift | NM_001126116 | M1fs | Chr17 | 7578532 | */-T | tAML-03 |
| *TP53BP1* | Nonsense | NM_001141980 | L807X | Chr15 | 43748386 | T>G | MDS-07 |
| *UBE2Q1* | Frameshift | NM_017582 | N412fs | Chr1 | 154523415 | */+A | MDS-12 |
| *ZRSR2* | Frameshift | NM_005089 | G323fs | ChrX | 15840883 | */-G | MDS-08 |
| *ZRSR2* | Nonsense | NM_005089 | E362X | ChrX | 15841000 | G>T | MDS-16 |

**Supplementary Table 4. Spliceosome mutations in myeloid neoplasms**

| | N | *U2AF35* | *SRSF2* | *ZRSR2* | *SF3B1* | *SF3A1* | *PRPF40B* | *U2AF65* | *SF1* | *SRSF1* | Total# |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MDS | 228 | 20 ( 8.8% ) | 23 ( 10.1% ) | 13 ( 5.7% ) | 67 (29.4% ) | 3 ( 1.3% ) | 3 ( 1.3% ) | 2 ( 0.9% ) | 3 ( 1.3% ) | 0 ( 0.0% ) | 130 ( 57.0% ) |
| RA | 8 | 1 ( 12.5% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 1 ( 12.5% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 2 ( 25.0% ) |
| RARS | 23 | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 19 ( 82.6% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 19 ( 82.6% ) |
| RCMD-RS | 50 | 0 ( 0.0% ) | 5 ( 10.0% ) | 1 ( 2.0% ) | 38 ( 76.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 1 ( 2.0% ) | 0 ( 0.0% ) | 43 ( 86.0% ) |
| RCMD | 39 | 7 ( 17.9% ) | 5 ( 12.8% ) | 3 ( 7.7% ) | 0 ( 0.0% ) | 1 ( 2.6% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 16 ( 41.0% ) |
| del (5q) | 8 | 0 ( 0.0% ) | 0 ( 0.0% ) | 1 ( 12.5% ) | 2 ( 25.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 1 ( 12.5% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 4 ( 50.0% ) |
| RAEB-1 | 45 | 8 ( 17.8% ) | 4 ( 8.9% ) | 5 ( 11.1% ) | 3 ( 6.7% ) | 0 ( 0.0% ) | 1 ( 2.2% ) | 1 ( 2.2% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 21 ( 46.7% ) |
| RAEB-2 | 46 | 4 ( 8.7% ) | 7 ( 15.2% ) | 2 ( 4.3% ) | 4 ( 8.7% ) | 2 ( 4.3% ) | 1 ( 2.2% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 19 ( 41.3% ) |
| MDS-U | 2 | 0 ( 0.0% ) | 1 ( 50.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 1 ( 50.0% ) | 0 ( 0.0% ) | 2 ( 100.0% ) |
| Unknown | 7 | 0 ( 0.0% ) | 1 ( 14.3% ) | 1 ( 14.3% ) | 1 ( 14.3% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 1 ( 14.3% ) | 0 ( 0.0% ) | 4 ( 57.1% ) |
| CMML | 88 | 7 ( 8.0% ) | 27 ( 30.7% ) | 8 ( 9.1% ) | 4 ( 4.5% ) | 2 ( 2.3% ) | 1 ( 1.1% ) | 2 ( 2.3% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 48 ( 54.5% ) |
| AML | 213 | 9 ( 4.2% ) | 5 ( 2.3% ) | 1 ( 0.5% ) | 7 ( 3.3% ) | 2 ( 0.9% ) | 2 ( 0.9% ) | 0 ( 0.0% ) | 1 ( 0.5% ) | 0 ( 0.0% ) | 26 ( 12.2% ) |
| *de novo* AML | 151 | 3 ( 2.0% ) | 1 ( 0.7% ) | 0 ( 0.0% ) | 4 ( 2.6% ) | 1 ( 0.7% ) | 1 ( 0.7% ) | 0 ( 0.0% ) | 1 ( 0.7% ) | 0 ( 0.0% ) | 10 ( 6.6% ) |
| other AML | 62 | 6 ( 9.7% ) | 4 ( 6.5% ) | 1 ( 1.6% ) | 3 ( 4.8% ) | 1 ( 1.6% ) | 1 ( 1.6% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 16 ( 25.8% ) |
| MPN | 53 | 1 ( 1.9% ) | 1 ( 1.9% ) | 1 ( 1.9% ) | 0 ( 0.0% ) | 0 ( 0.0% ) | 1 ( 1.9% ) | 0 ( 0.0% ) | 1 ( 1.9% ) | 0 ( 0.0% ) | 5 ( 9.4% ) |
| Total | 582 | 37 ( 6.4% ) | 56 ( 9.6% ) | 23 ( 4.0% ) | 78 ( 13.4% ) | 7 ( 1.2% ) | 7 ( 1.2% ) | 4 ( 0.7% ) | 5 ( 0.9% ) | 0 ( 0.0% ) | 209 ( 35.9% ) |

Abbreviations: RA, refractory anemia;RCMD, refractory cytopenia with multilineage dysplasia;
del(5q), MDS with isolated del(5q); RAEB-1, refractory anemia with excess blasts-1;
RAEB-2, refractory anemia with excess blasts-2;MDS-U, MDS unclassifiable;
Other AML includes acute myeloid leukeimia with myelodysplasia-related changes and therapy-related myeloid neoplasms.
#Cases with mutaions of more than two genes were treated as single cases in total.

Supplementary Table 5. Mutation analysis of spliceosome mutaions in 582 cases with myeloid neoplasms

| Sample ID | Diagnosis | IPSS score | Spiceosome Mutation | Mutaion Type | RefSeq | Chr | Position | Allele Change | Prediction by MutationTaster[#] |
|---|---|---|---|---|---|---|---|---|---|
| MDS-001 | RA | 0.5 | – | – | – | – | – | – | – |
| MDS-002 | RA | 0.5 | – | – | – | – | – | – | – |
| MDS-003 | RA | 0.5 | – | – | – | – | – | – | – |
| MDS-004 | RA | 0 | U2AF35_Q157R | Missense | NM_006758 | Chr21 | 44514777 | A>G | Disease_causing |
| MDS-005 | RA | 0 | – | – | – | – | – | – | – |
| MDS-006 | RA | ND | – | – | – | – | – | – | – |
| MDS-007 | RA | 0.5 | PRPF40B_P212L | Missense | NM_001031698 | Chr12 | 50027764 | C>T | Polymorphism |
| MDS-008 | RA | 0 | – | – | – | – | – | – | – |
| MDS-009 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-010 | RARS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-011 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-012 | RARS | 0 | – | – | – | – | – | – | – |
| MDS-013 | RARS | 0 | SF3B1_K666E | Missense | NM_012433 | Chr2 | 198267361 | A>G | Disease_causing |
| MDS-014 | RARS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-015 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-016 | RARS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-017 | RARS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>A | Disease_causing |
| MDS-018 | RARS | 0 | – | – | – | – | – | – | – |
| MDS-019 | RARS | ND | – | – | – | – | – | – | – |
| MDS-020 | RARS | 0.5 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-021 | RARS | 0 | SF3B1_R625L | Missense | NM_012433 | Chr2 | 198267483 | G>T | Disease_causing |
| MDS-022 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-023 | RARS | 0.5 | – | – | – | – | – | – | – |
| MDS-024 | RARS | 0.5 | SF3B1_K666R | Missense | NM_012433 | Chr2 | 198267360 | A>G | Disease_causing |
| MDS-025 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-026 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-027 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-028 | RARS | 0 | SF3B1_E622D | Missense | NM_012433 | Chr2 | 198267491 | G>T | Disease_causing |
| MDS-029 | RARS | 0.5 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-030 | RARS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-031 | RARS | 1 | SF3B1_R625C | Missense | NM_012433 | Chr2 | 198267484 | C>T | Disease_causing |
| MDS-032 | RCMD-RS | 0.5 | – | – | – | – | – | – | – |
| MDS-033 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-034 | RCMD-RS | 0 | SF3B1_E622D | Missense | NM_012433 | Chr2 | 198267491 | G>T | Disease_causing |
| MDS-035 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-036 | RCMD-RS | 0 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| MDS-037 | RCMD-RS | ND | SF3B1_D781G | Missense | NM_012433 | Chr2 | 198266494 | A>G | Disease_causing |
| MDS-038 | RCMD-RS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>G | Disease_causing |
| MDS-039 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-040 | RCMD-RS | 0 | SF3B1_R625L | Missense | NM_012433 | Chr2 | 198267483 | G>T | Disease_causing |
| MDS-041 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-042 | RCMD-RS | 0.5 | – | – | – | – | – | – | – |
| MDS-043 | RCMD-RS | 0 | – | – | – | – | – | – | – |
| MDS-044 | RCMD-RS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>G | Disease_causing |
| MDS-045 | RCMD-RS | 0 | SF3B1_E622D | Missense | NM_012433 | Chr2 | 198267491 | G>C | Disease_causing |
| MDS-046 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-047 | RCMD-RS | 0 | SF3B1_H662D | Missense | NM_012433 | Chr2 | 198267373 | C>G | Disease_causing |
| MDS-048 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-049 | RCMD-RS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>G | Disease_causing |
| MDS-050 | RCMD-RS | 0 | SF3B1_N626D | Missense | NM_012433 | Chr2 | 198267481 | A>G | Disease_causing |
| MDS-051 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-052 | RCMD-RS | 0 | SF3B1_K666R | Missense | NM_012433 | Chr2 | 198267360 | A>G | Disease_causing |
| MDS-053 | RCMD-RS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>A | Disease_causing |
| MDS-054 | RCMD-RS | 0 | – | – | – | – | – | – | – |
| MDS-055 | RCMD-RS | 0.5 | ZRSR2_Y347X | Nonsense | NM_005089 | ChrX | 15840957 | C>G | Disease_causing |
| MDS-056 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-057 | RCMD-RS | ND | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-058 | RCMD-RS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>G | Disease_causing |
| MDS-058 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-059 | RCMD-RS | ND | SF3B1_K666T | Missense | NM_012433 | Chr2 | 198267360 | A>C | Disease_causing |
| MDS-060 | RCMD-RS | 0 | SF3B1_K666N | Missense | NM_012433 | Chr2 | 198267359 | G>T | Disease_causing |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| MDS-060 | RCMD-RS | 0 | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| MDS-061 | RCMD-RS | ND | SF1_T474A | Missense | NM_201997 | Chr11 | 64534538 | A>G | Polymorphism |
| MDS-061 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-062 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-063 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-064 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-065 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-066 | RCMD-RS | 0 | SF3B1_K666E | Missense | NM_012433 | Chr2 | 198267361 | A>G | Disease_causing |
| MDS-067 | RCMD-RS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>A | Disease_causing |
| MDS-068 | RCMD-RS | 0 | SF3B1_H662D | Missense | NM_012433 | Chr2 | 198267373 | C>G | Disease_causing |
| MDS-069 | RCMD-RS | 0 | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| MDS-070 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-071 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-072 | RCMD-RS | 1 | – | – | – | – | – | – | – |
| MDS-073 | RCMD-RS | ND | SF3B1_E622D | Missense | NM_012433 | Chr2 | 198267491 | G>T | Disease_causing |
| MDS-074 | RCMD-RS | ND | SF3B1_Y623C | Missense | NM_012433 | Chr2 | 198267489 | A>G | Disease_causing |
| MDS-075 | RCMD-RS | ND | – | – | – | – | – | – | – |
| MDS-076 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-077 | RCMD-RS | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-078 | RCMD-RS | 0 | SF3B1_H662Q | Missense | NM_012433 | Chr2 | 198267371 | C>G | Disease_causing |
| MDS-079 | RCMD-RS | 0 | – | – | – | – | – | – | – |
| MDS-080 | RCMD-RS | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-081 | RCMD-RS | 0 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-082 | RCMD | 0.5 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-083 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-084 | RCMD | ND | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-085 | RCMD | 0 | U2AF35_Q157P | Missense | NM_006758 | Chr21 | 44514777 | A>C | Disease_causing |
| MDS-086 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-087 | RCMD | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-088 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-089 | RCMD | 0 | – | – | – | – | – | – | – |
| MDS-090 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-091 | RCMD | ND | ZRSR2_C302R | Missense | NM_005089 | ChrX | 15838406 | T>C | Disease_causing |
| MDS-092 | RCMD | ND | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-093 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-094 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-095 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-096 | RCMD | ND | – | – | – | – | – | – | – |
| MDS-097 | RCMD | 0 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-098 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-099 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-100 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-101 | RCMD | 1 | – | – | – | – | – | – | – |
| MDS-102 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-103 | RCMD | 1 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-104 | RCMD | 0.5 | U2AF35_Q157R | Missense | NM_006758 | Chr21 | 44514777 | A>G | Disease_causing |
| MDS-105 | RCMD | 0.5 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-106 | RCMD | ND | – | – | – | – | – | – | – |
| MDS-107 | RCMD | 1 | SF3A1_K166T | Missense | NM_005877 | Chr22 | 30738828 | A>C | Disease_causing |
| MDS-108 | RCMD | 0.5 | ZRSR2_G323fs | Frameshift | NM_005089 | ChrX | 15840883 | */-G | Disease_causing |
| MDS-109 | RCMD | ND | ZRSR2_K257_splice | Splice site | NM_005089 | ChrX | 15834014 | G>A | Disease_causing |
| MDS-110 | RCMD | ND | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| MDS-111 | RCMD | ND | U2AF35_A26V | Missense | NM_006758 | Chr21 | 44524480 | C>T | Disease_causing |
| MDS-112 | RCMD | 1 | – | – | – | – | – | – | – |
| MDS-113 | RCMD | 1 | – | – | – | – | – | – | – |
| MDS-114 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-115 | RCMD | ND | U2AF35_Q157P | Missense | NM_006758 | Chr21 | 44514777 | A>C | Disease_causing |
| MDS-116 | RCMD | 0 | – | – | – | – | – | – | – |
| MDS-117 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-118 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-119 | RCMD | 0.5 | – | – | – | – | – | – | – |
| MDS-120 | RCMD | 1.5 | – | – | – | – | – | – | – |
| MDS-121 | del(5q) | ND | U2AF65_R18W | Missense | NM_007279 | Chr19 | 56170578 | C>T | Disease_causing |
| MDS-122 | del(5q) | ND | ZRSR2_I53T | Missense | NM_005089 | ChrX | 15818031 | T>C | Polymorphism |
| MDS-123 | del(5q) | ND | – | – | – | – | – | – | – |
| MDS-124 | del(5q) | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| MDS-125 | del(5q) | ND | SF3B1_K182E | Missense | NM_012433 | Chr2 | 198281587 | A>G | Disease_causing |
| MDS-126 | del(5q) | ND | – | – | – | – | – | – | – |
| MDS-127 | del(5q) | ND | – | – | – | – | – | – | – |
| MDS-128 | del(5q) | ND | – | – | – | – | – | – | – |
| MDS-129 | RAEB-1 | 0.5 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-130 | RAEB-1 | 0.5 | – | – | – | – | – | – | – |
| MDS-131 | RAEB-1 | ND | – | – | – | – | – | – | – |
| MDS-132 | RAEB-1 | 0.5 | – | – | – | – | – | – | – |
| MDS-133 | RAEB-1 | ND | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| MDS-134 | RAEB-1 | 1 | U2AF35_Q157R | Missense | NM_006758 | Chr21 | 44514777 | A>G | Disease_causing |
| MDS-135 | RAEB-1 | 1 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-136 | RAEB-1 | 0.5 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-136 | RAEB-1 | 0.5 | U2AF65_M144I | Missense | NM_007279 | Chr19 | 56172501 | G>A | Disease_causing |
| MDS-137 | RAEB-1 | 0.5 | SF3B1_K666N | Missense | NM_012433 | Chr2 | 198267359 | G>T | Disease_causing |
| MDS-138 | RAEB-1 | ND | – | – | – | – | – | – | – |
| MDS-139 | RAEB-1 | 1 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-140 | RAEB-1 | 1.5 | – | – | – | – | – | – | – |
| MDS-141 | RAEB-1 | 3 | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| MDS-142 | RAEB-1 | 1 | SF3B1_K666T | Missense | NM_012433 | Chr2 | 198267360 | A>C | Disease_causing |
| MDS-143 | RAEB-1 | 1.5 | – | – | – | – | – | – | – |
| MDS-144 | RAEB-1 | 1 | – | – | – | – | – | – | – |
| MDS-145 | RAEB-1 | 1.5 | – | – | – | – | – | – | – |
| MDS-146 | RAEB-1 | 0.5 | – | – | – | – | – | – | – |
| MDS-147 | RAEB-1 | 0.5 | ZRSR2_N382K | Missense | NM_005089 | Chr19 | 56172501 | G>A | Polymorphism |
| MDS-148 | RAEB-1 | 1 | ZRSR2_E118fs | Frameshift | NM_005089 | ChrX | 15822272 | */-GA | Disease_causing |
| MDS-149 | RAEB-1 | ND | PRPF40B_P15H | Missense | NM_001031698 | Chr12 | 50025209 | C>A | Polymorphism |
| MDS-150 | RAEB-1 | ND | SF3B1_K666N | Missense | NM_012433 | Chr2 | 198267359 | G>T | Disease_causing |
| MDS-151 | RAEB-1 | 1 | – | – | – | – | – | – | – |
| MDS-152 | RAEB-1 | ND | ZRSR2_S40X | Nonsense | NM_005089 | ChrX | 15809134 | C>G | Disease_causing |
| MDS-153 | RAEB-1 | 2 | – | – | – | – | – | – | – |
| MDS-154 | RAEB-1 | 0.5 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-155 | RAEB-1 | 1.5 | – | – | – | – | – | – | – |
| MDS-156 | RAEB-1 | 1 | ZRSR2_E362X | Nonsense | NM_005089 | ChrX | 15841000 | G>T | Disease_causing |
| MDS-157 | RAEB-1 | 0.5 | U2AF35_Q157R | Missense | NM_006758 | Chr21 | 44514777 | A>G | Disease_causing |
| MDS-158 | RAEB-1 | 0.5 | – | – | – | – | – | – | – |
| MDS-159 | RAEB-1 | 2 | – | – | – | – | – | – | – |
| MDS-160 | RAEB-1 | ND | – | – | – | – | – | – | – |
| MDS-161 | RAEB-1 | ND | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-162 | RAEB-1 | ND | – | – | – | – | – | – | – |
| MDS-163 | RAEB-1 | ND | – | – | – | – | – | – | – |
| MDS-164 | RAEB-1 | ND | ZRSR2_E133G | Missense | NM_005089 | ChrX | 15822319 | A>G | Disease_causing |
| MDS-165 | RAEB-1 | 2.5 | – | – | – | – | – | – | – |
| MDS-166 | RAEB-1 | 0 | – | – | – | – | – | – | – |
| MDS-167 | RAEB-1 | 0 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-168 | RAEB-1 | 3 | – | – | – | – | – | – | – |
| MDS-169 | RAEB-1 | 2 | U2AF35_S34Y | Missense | NM_006758 | Chr21 | 44524456 | C>A | Disease_causing |
| MDS-170 | RAEB-1 | 2 | – | – | – | – | – | – | – |
| MDS-171 | RAEB-1 | 2 | – | – | – | – | – | – | – |
| MDS-172 | RAEB-1 | 2.5 | – | – | – | – | – | – | – |
| MDS-173 | RAEB-1 | 1.5 | – | – | – | – | – | – | – |
| MDS-174 | RAEB-2 | 2.5 | – | – | – | – | – | – | – |
| MDS-175 | RAEB-2 | 1.5 | – | – | – | – | – | – | – |
| MDS-176 | RAEB-2 | 2 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-177 | RAEB-2 | 2 | – | – | – | – | – | – | – |
| MDS-178 | RAEB-2 | 2 | SF3A1_M667V | Missense | NM_005877 | Chr22 | 30733122 | A>G | Disease_causing |
| MDS-178 | RAEB-2 | 2 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-179 | RAEB-2 | 1.5 | SF3B1_K666N | Missense | NM_012433 | Chr2 | 198267359 | G>T | Disease_causing |
| MDS-180 | RAEB-2 | ND | – | – | – | – | – | – | – |
| MDS-181 | RAEB-2 | ND | – | – | – | – | – | – | – |
| MDS-182 | RAEB-2 | 0.5 | PRPF40B_P540S | Missense | NM_001031698 | Chr12 | 50031524 | C>T | Disease_causing |
| MDS-183 | RAEB-2 | 2.5 | – | – | – | – | – | – | – |
| MDS-184 | RAEB-2 | 0.5 | – | – | – | – | – | – | – |
| MDS-185 | RAEB-2 | 3 | – | – | – | – | – | – | – |
| MDS-186 | RAEB-2 | ND | – | – | – | – | – | – | – |
| MDS-187 | RAEB-2 | 2 | – | – | – | – | – | – | – |
| MDS-188 | RAEB-2 | 2 | – | – | – | – | – | – | – |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| MDS-189 | RAEB-2 | 2 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-190 | RAEB-2 | 2.5 | – | – | – | – | – | – | – |
| MDS-191 | RAEB-2 | 3 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| MDS-192 | RAEB-2 | 2 | ZRSR2_L237fs | Frameshift | NM_005089 | ChrX | 15833950 | */-C | Disease_causing |
| MDS-193 | RAEB-2 | 2.5 | – | – | – | – | – | – | – |
| MDS-194 | RAEB-2 | 2.5 | – | – | – | – | – | – | – |
| MDS-195 | RAEB-2 | 2.5 | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| MDS-196 | RAEB-2 | 2.5 | – | – | – | – | – | – | – |
| MDS-197 | RAEB-2 | 3.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-198 | RAEB-2 | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-199 | RAEB-2 | 2 | – | – | – | – | – | – | – |
| MDS-200 | RAEB-2 | 2 | – | – | – | – | – | – | – |
| MDS-201 | RAEB-2 | 2 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| MDS-202 | RAEB-2 | 1.5 | – | – | – | – | – | – | – |
| MDS-203 | RAEB-2 | 1.5 | U2AF35_Q157P | Missense | NM_006758 | Chr21 | 44514777 | A>C | Disease_causing |
| MDS-204 | RAEB-2 | 2 | – | – | – | – | – | – | – |
| MDS-205 | RAEB-2 | 1.5 | – | – | – | – | – | – | – |
| MDS-206 | RAEB-2 | 2 | – | – | – | – | – | – | – |
| MDS-207 | RAEB-2 | 2.5 | SF3A1_A57S | Missense | NM_005877 | Chr22 | 30748956 | G>T | Disease_causing |
| MDS-208 | RAEB-2 | 2.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-209 | RAEB-2 | ND | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| MDS-210 | RAEB-2 | 1.5 | U2AF35_S34Y | Missense | NM_006758 | Chr21 | 44524456 | C>A | Disease_causing |
| MDS-211 | RAEB-2 | ND | – | – | – | – | – | – | – |
| MDS-212 | RAEB-2 | 2.5 | ZRSR2_R68_splice | Splice site | NM_005089 | ChrX | 15818078 | T>G | Disease_causing |
| MDS-213 | RAEB-2 | 1.5 | – | – | – | – | – | – | – |
| MDS-214 | RAEB-2 | 1.5 | – | – | – | – | – | – | – |
| MDS-215 | RAEB-2 | 2 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| MDS-216 | RAEB-2 | 2 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MDS-217 | RAEB-2 | 2 | – | – | – | – | – | – | – |
| MDS-218 | RAEB-2 | ND | – | – | – | – | – | – | – |
| MDS-219 | RAEB-2 | ND | – | – | – | – | – | – | – |
| MDS-220 | MDS-U | 0 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| MDS-221 | MDS-U | 0.5 | SF1_A508G | Missense | NM_201997 | Chr11 | 64534435 | C>G | Polymorphism |
| MDS-222 | MDS | ND | SF3B1_K666N | Missense | NM_012433 | Chr2 | 198267359 | G>C | Disease_causing |
| MDS-223 | MDS | ND | SF1_G372V | Missense | NM_201997 | Chr11 | 64535270 | G>T | Disease_causing |
| MDS-224 | MDS | ND | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| MDS-225 | MDS | ND | ZRSR2_A96fs | Frameshift | NM_005089 | ChrX | 15821895 | */+A | Disease_causing |
| MDS-226 | MDS | 0.5 | – | – | – | – | – | – | – |
| MDS-227 | MDS | ND | – | – | – | – | – | – | – |
| MDS-228 | MDS | ND | – | – | – | – | – | – | – |
| CMML-001 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-002 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-003 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-004 | CMML1 | 1 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| CMML-005 | CMML1 | ND | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| CMML-006 | CMML1 | ND | U2AF65_L187V | Missense | NM_007279 | Chr19 | 56173940 | T>G | Disease_causing |
| CMML-007 | CMML1 | 1 | – | – | – | – | – | – | – |
| CMML-008 | CMML1 | 1.5 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| CMML-008 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-009 | CMML1 | 0 | – | – | – | – | – | – | – |
| CMML-010 | CMML1 | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| CMML-011 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-012 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-013 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-014 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-015 | CMML1 | 1 | ZRSR2_A96fs | Frameshift | NM_005089 | ChrX | 15821895 | */+A | Disease_causing |
| CMML-016 | CMML1 | 2 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| CMML-017 | CMML1 | 1 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| CMML-018 | CMML1 | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-019 | CMML1 | 0.5 | U2AF35_Q157P | Missense | NM_006758 | Chr21 | 44514777 | A>C | Disease_causing |
| CMML-020 | CMML1 | 0.5 | PRPF40B_M58I | Missense | NM_001031698 | Chr12 | 50025654 | G>A | Polymorphism |
| CMML-021 | CMML1 | 0.5 | ZRSR2_I202N | Missense | NM_005089 | ChrX | 15833847 | T>A | Disease_causing |
| CMML-022 | CMML1 | 0 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-023 | CMML1 | 0.5 | ZRSR2_H330R | Missense | NM_005089 | ChrX | 15840905 | A>G | Disease_causing |
| CMML-023 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-024 | CMML1 | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| CMML-025 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-026 | CMML1 | 1.5 | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| CMML-027 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-028 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-029 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-030 | CMML1 | ND | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-031 | CMML1 | 0 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-032 | CMML1 | 0 | U2AF65_L187V | Missense | NM_007279 | Chr19 | 56173940 | T>G | Disease_causing |
| CMML-033 | CMML1 | 1.5 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| CMML-034 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-035 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-036 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-037 | CMML1 | 0.5 | ZRSR2_N327fs | Frameshift | NM_005089 | ChrX | 15840897 | */+T | Disease_causing |
| CMML-038 | CMML1 | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-039 | CMML1 | ND | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| CMML-040 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-041 | CMML1 | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-042 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-043 | CMML1 | ND | – | – | – | – | – | – | – |
| CMML-044 | CMML1 | 0 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-045 | CMML1 | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-046 | CMML1 | ND | U2AF35_Q157R | Missense | NM_006758 | Chr21 | 44514777 | A>G | Disease_causing |
| CMML-047 | CMML1 | 0.5 | – | – | – | – | – | – | – |
| CMML-048 | CMML1 | 0.5 | ZRSR2_F239V | Missense | NM_005089 | ChrX | 15833957 | T>G | Disease_causing |
| CMML-049 | CMML1 | ND | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-050 | CMML1 | 1.5 | – | – | – | – | – | – | – |
| CMML-051 | CMML1 | 0 | – | – | – | – | – | – | – |
| CMML-052 | CMML1 | 0.5 | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| CMML-053 | CMML1 | 0.5 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| CMML-054 | CMML1 | 0.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-055 | CMML2 | ND | – | – | – | – | – | – | – |
| CMML-056 | CMML2 | ND | – | – | – | – | – | – | – |
| CMML-057 | CMML2 | ND | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-058 | CMML2 | 2 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| CMML-059 | CMML2 | ND | – | – | – | – | – | – | – |
| CMML-060 | CMML2 | 2 | – | – | – | – | – | – | – |
| CMML-061 | CMML2 | ND | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-062 | CMML2 | ND | – | – | – | – | – | – | – |
| CMML-063 | CMML2 | 2 | – | – | – | – | – | – | – |
| CMML-064 | CMML2 | 2.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-065 | CMML2 | 2.5 | – | – | – | – | – | – | – |
| CMML-066 | CMML2 | 3.5 | – | – | – | – | – | – | – |
| CMML-067 | CMML2 | 3 | – | – | – | – | – | – | – |
| CMML-068 | CMML2 | 0.5 | ZRSR2_R126X | Nonsense | NM_005089 | ChrX | 15822297 | C>T | Disease_causing |
| CMML-069 | CMML2 | 3 | – | – | – | – | – | – | – |
| CMML-070 | CMML2 | 2.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-071 | CMML2 | 2.5 | – | – | – | – | – | – | – |
| CMML-072 | CMML2 | 2 | ZRSR2_W291X | Nonsense | NM_005089 | ChrX | 15838374 | G>A | Disease_causing |
| CMML-073 | CMML2 | 2 | – | – | – | – | – | – | – |
| CMML-074 | CMML2 | ND | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| CMML-075 | CMML2 | 2 | ZRSR2_E148X | Nonsense | NM_005089 | ChrX | 15827326 | G>T | Disease_causing |
| CMML-076 | CMML2 | 2 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| CMML-077 | CMML2 | 1.5 | SF3A1_I141M | Missense | NM_005877 | Chr22 | 30741150 | C>G | Polymorphism |
| CMML-078 | CMML2 | ND | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| CMML-079 | CMML2 | 2 | – | – | – | – | – | – | – |
| CMML-080 | CMML2 | ND | – | – | – | – | – | – | – |
| CMML-081 | CMML2 | 2.5 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-082 | CMML2 | 1.5 | – | – | – | – | – | – | – |
| CMML-083 | CMML2 | ND | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| CMML-084 | CMML2 | 1.5 | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| CMML-085 | CMML2 | 2.5 | – | – | – | – | – | – | – |
| CMML-086 | CMML2 | ND | SF3A1_E373D | Missense | NM_005877 | Chr22 | 30736754 | G>T | Disease_causing |
| CMML-086 | CMML2 | ND | SF3A1_T374P | Missense | NM_005877 | Chr22 | 30736753 | A>C | Disease_causing |
| CMML-086 | CMML2 | ND | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| CMML-087 | CMML2 | 1 | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| CMML-088 | CMML2 | 2 | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| AML-001 | AML M0 | NA | – | – | – | – | – | – | – |
| AML-002 | AML M0 | NA | – | – | – | – | – | – | – |
| AML-003 | AML M0 | NA | – | – | – | – | – | – | – |
| AML-004 | AML M0 | NA | – | – | – | – | – | – | – |
| AML-005 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-006 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-007 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-008 | AML M1 | NA | SF1_Y476C | Missense | NM_201997 | Chr11 | 64534531 | A>G | Polymorphism |
| AML-008 | AML M1 | NA | U2AF35_Q157R | Missense | NM_006758 | Chr21 | 44514777 | A>G | Disease_causing |
| AML-009 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-010 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-011 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-012 | AML, M1 | NA | – | – | – | – | – | – | – |
| AML-013 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-014 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-015 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-016 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-017 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-018 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-019 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-020 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-021 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-022 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-023 | AML M1 | NA | – | – | – | – | – | – | – |
| AML-024 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-025 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-026 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-027 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-028 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-029 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-030 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-031 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-032 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-033 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-034 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-035 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-036 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-037 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-038 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-039 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-040 | AML M2 | NA | SF3B1_G347V | Missense | NM_012433 | Chr2 | 198273170 | G>T | Disease_causing |
| AML-041 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-042 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-043 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-044 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-045 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-046 | AML M2 | NA | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| AML-047 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-048 | AML M2 | NA | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| AML-049 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-050 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-051 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-052 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-053 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-054 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-055 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-056 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-057 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-058 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-059 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-060 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-061 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-062 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-063 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-064 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-065 | AML M2 | NA | – | – | – | – | – | – | – |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| AML-066 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-067 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-068 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-069 | AML M2 | NA | SRSF2_P95L | Missense | NM_003016 | Chr17 | 74732959 | C>T | Disease_causing |
| AML-070 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-071 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-072 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-073 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-074 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-075 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-076 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-077 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-078 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-079 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-080 | AML M2 | NA | – | – | – | – | – | – | – |
| AML-081 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-082 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-083 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-084 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-085 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-086 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-087 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-088 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-089 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-090 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-091 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-092 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-093 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-094 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-095 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-096 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-097 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-098 | AML M3 | NA | PRPF40B_P540S | Missense | NM_001031698 | Chr12 | 50031524 | C>T | Disease_causing |
| AML-099 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-100 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-101 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-102 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-103 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-104 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-105 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-106 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-107 | AML M3 | NA | – | – | – | – | – | – | – |
| AML-108 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-109 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-110 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-111 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-112 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-113 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-114 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-115 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-116 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-117 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-118 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-119 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-120 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-121 | AML M4 | NA | – | – | – | – | – | – | – |
| AML-122 | AML M4 | NA | SF3A1_Y772C | Missense | NM_005877 | Chr22 | 30730650 | A>G | Disease_causing |
| AML-123 | AML M4 | NA | SF3B1_K666N | Missense | NM_012433 | Chr2 | 198267359 | G>T | Disease_causing |
| AML-124 | AML M4 | NA | SF3B1_K666T | Missense | NM_012433 | Chr2 | 198267360 | A>C | Disease_causing |
| AML-125 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-126 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-127 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-128 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-129 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-130 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-131 | AML M5 | NA | – | – | – | – | – | – | – |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| AML-132 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-133 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-134 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-135 | AML M5 | NA | – | – | – | – | – | – | – |
| AML-136 | AML M6 | NA | – | – | – | – | – | – | – |
| AML-137 | AML M6 | NA | – | – | – | – | – | – | – |
| AML-138 | AML M6 | NA | – | – | – | – | – | – | – |
| AML-139 | AML M6 | NA | – | – | – | – | – | – | – |
| AML-140 | AML M6 | NA | – | – | – | – | – | – | – |
| AML-141 | AML M7 | NA | – | – | – | – | – | – | – |
| AML-142 | AML | NA | – | – | – | – | – | – | – |
| AML-143 | AML | NA | – | – | – | – | – | – | – |
| AML-144 | AML | NA | U2AF35_S34Y | Missense | NM_006758 | Chr21 | 44524456 | C>A | Disease_causing |
| AML-145 | AML | NA | – | – | – | – | – | – | – |
| AML-146 | AML | NA | – | – | – | – | – | – | – |
| AML-147 | AML | NA | – | – | – | – | – | – | – |
| AML-148 | AML | NA | – | – | – | – | – | – | – |
| AML-149 | AML | NA | – | – | – | – | – | – | – |
| AML-150 | AML | NA | – | – | – | – | – | – | – |
| AML-151 | AML | NA | – | – | – | – | – | – | – |
| MPN-001 | ET | NA | – | – | – | – | – | – | – |
| MPN-002 | ET | NA | – | – | – | – | – | – | – |
| MPN-003 | ET | NA | – | – | – | – | – | – | – |
| MPN-004 | ET | NA | – | – | – | – | – | – | – |
| MPN-005 | ET | NA | PRPF40B_D442N | Missense | NM_001031698 | Chr12 | 50029740 | G>A | Disease_causing |
| MPN-006 | ET | NA | – | – | – | – | – | – | – |
| MPN-007 | ET | NA | – | – | – | – | – | – | – |
| MPN-008 | ET | NA | – | – | – | – | – | – | – |
| MPN-009 | ET | NA | – | – | – | – | – | – | – |
| MPN-010 | ET | NA | – | – | – | – | – | – | – |
| MPN-011 | ET | NA | – | – | – | – | – | – | – |
| MPN-012 | ET | NA | – | – | – | – | – | – | – |
| MPN-013 | ET | NA | – | – | – | – | – | – | – |
| MPN-014 | ET | NA | – | – | – | – | – | – | – |
| MPN-015 | ET | NA | – | – | – | – | – | – | – |
| MPN-016 | ET | NA | – | – | – | – | – | – | – |
| MPN-017 | ET | NA | – | – | – | – | – | – | – |
| MPN-018 | ET | NA | – | – | – | – | – | – | – |
| MPN-019 | ET | NA | ZRSR2_C326R | Missense | NM_005089 | ChrX | 15840892 | T>C | Disease_causing |
| MPN-020 | ET | NA | – | – | – | – | – | – | – |
| MPN-021 | ET | NA | – | – | – | – | – | – | – |
| MPN-022 | ET | NA | – | – | – | – | – | – | – |
| MPN-023 | ET | NA | – | – | – | – | – | – | – |
| MPN-024 | ET | NA | – | – | – | – | – | – | – |
| MPN-025 | ET | NA | – | – | – | – | – | – | – |
| MPN-026 | ET | NA | – | – | – | – | – | – | – |
| MPN-027 | ET | NA | – | – | – | – | – | – | – |
| MPN-028 | ET | NA | – | – | – | – | – | – | – |
| MPN-029 | PMF | NA | – | – | – | – | – | – | – |
| MPN-030 | PMF | NA | – | – | – | – | – | – | – |
| MPN-031 | PMF | NA | – | – | – | – | – | – | – |
| MPN-032 | PMF | NA | – | – | – | – | – | – | – |
| MPN-033 | PMF | NA | SF1_T454M | Missense | NM_201997 | Chr11 | 64534705 | C>T | Polymorphism |
| MPN-034 | PMF | NA | – | – | – | – | – | – | – |
| MPN-035 | PMF | NA | SRSF2_P95H | Missense | NM_003016 | Chr17 | 74732959 | C>A | Disease_causing |
| MPN-036 | PMF | NA | – | – | – | – | – | – | – |
| MPN-037 | PMF | NA | – | – | – | – | – | – | – |
| MPN-038 | PMF | NA | – | – | – | – | – | – | – |
| MPN-039 | PMF | NA | – | – | – | – | – | – | – |
| MPN-040 | PMF | NA | – | – | – | – | – | – | – |
| MPN-041 | PMF | NA | U2AF35_S34Y | Missense | NM_006758 | Chr21 | 44524456 | C>A | Disease_causing |
| MPN-042 | PMF | NA | – | – | – | – | – | – | – |
| MPN-043 | PV | NA | – | – | – | – | – | – | – |
| MPN-044 | PV | NA | – | – | – | – | – | – | – |
| MPN-045 | PV | NA | – | – | – | – | – | – | – |
| MPN-046 | PV | NA | – | – | – | – | – | – | – |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| MPN-047 | PV | NA | – | – | – | – | – | – | – |
| MPN-048 | PV | NA | – | – | – | – | – | – | – |
| MPN-049 | PV | NA | – | – | – | – | – | – | – |
| MPN-050 | PV | NA | – | – | – | – | – | – | – |
| MPN-051 | PV | NA | – | – | – | – | – | – | – |
| MPN-052 | MPN | NA | – | – | – | – | – | – | – |
| MPN-053 | MPN | NA | – | – | – | – | – | – | – |
| tAML-001 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-002 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-003 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-004 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-005 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-006 | AML/MRC | NA | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| tAML-007 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-008 | AML/MRC | NA | U2AF35_Q157R | Missense | NM_006758 | Chr21 | 44514777 | A>G | Disease_causing |
| tAML-009 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-010 | AML/MRC | NA | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| tAML-011 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-012 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-013 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-014 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-015 | AML/MRC | NA | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| tAML-016 | AML/MRC | NA | U2AF35_S34F | Missense | NM_006758 | Chr21 | 44524456 | C>T | Disease_causing |
| tAML-017 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-018 | AML/MRC | NA | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| tAML-019 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-020 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-021 | AML/MRC | NA | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| tAML-022 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-023 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-024 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-025 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-026 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-027 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-028 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-029 | AML/MRC | NA | PRPF40B_P383L | Missense | NM_001031698 | Chr12 | 50029195 | C>T | Disease_causing |
| tAML-030 | AML/MRC | NA | ZRSR2_N261Y | Missense | NM_005089 | ChrX | 15836719 | A>T | Disease_causing |
| tAML-031 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-032 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-033 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-034 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-035 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-036 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-037 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-038 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-039 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-040 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-041 | AML/MRC | NA | SF3A1_M117I | Missense | NM_005877 | Chr22 | 30742343 | G>C | NA |
| tAML-042 | AML/MRC | NA | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| tAML-043 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-044 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-045 | AML/MRC | NA | SF3B1_K700E | Missense | NM_012433 | Chr2 | 198266834 | A>G | Disease_causing |
| tAML-046 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-047 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-048 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-049 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-050 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-051 | AML/MRC | NA | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| tAML-052 | AML/MRC | NA | U2AF35_S34Y | Missense | NM_006758 | Chr21 | 44524456 | C>A | Disease_causing |
| tAML-053 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-054 | AML/MRC | NA | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| tAML-055 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-056 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-057 | AML/MRC | NA | SRSF2_P95R | Missense | NM_003016 | Chr17 | 74732959 | C>G | Disease_causing |
| tAML-058 | AML/MRC | NA | – | – | – | – | – | – | – |
| tAML-059 | Therapy-related AML | NA | – | – | – | – | – | – | – |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| tAML-060 | Therapy-related AML | NA | – | – | – | – | – | – | – |
| tAML-061 | Therapy-related AML | NA | – | – | – | – | – | – | – |
| tAML-062 | Therapy-related AML | NA | – | – | – | – | – | – | – |

*#Pathogeneity of somatic mutations was predicted using Mutation Taster (http://www.mutationtaster.org/)*

Abbreviations: IPSS, the International Prognostic Scoring System;

RA, refractory anemia; RCMD, refractory cytopenia with multilineage dysplasia;

del(5q), MDS with isolated del(5q); RAEB-1, refractory anemia with excess blasts-1;

RAEB-2, refractory anemia with excess blasts-2; MDS-U, MDS unclassifiable;
CMML1, chronic myelomonocytic leukemia 1; CMML2, chronic myelomonocytic leukemia 2;
ET, essential thrombocythemia; PMF, primary myelofibrosis; PV, polycythemia vera;
AML/MRC, acute myeloid leukemia with myelodysplasia related changes; ND, not determined
NA, not applicable

**Supplementary Table 6. The gene sets for RNA export from nucleus and extended NMD genes**

| geneset | RNA export from nucleus | NMD factors |
|---|---|---|
| size | 21 | 17 |
| | SMG5 * # | SMG5 |
| | UPF1 * # | SMG9 / FLJ12886 |
| | DDX19B * | DHX34 |
| | SMG1 * # | UPF1 |
| | NCBP2 * | SMG1 |
| | DDX39 * | UPF3B |
| | HRB * | BTZ / CASC3 |
| | UPF2 * # | UPF2 |
| | NUDT4 * | SMG7 |
| | SMG7 * # | MAGOH |
| | RAE1 * | SMG6 |
| | SMG6 * # | SMG8 / C17ORF71 |
| | TSC1 | Y14 / RBM8A |
| | BAT1 | eIF4A3 / DDX48 |
| | KHDRBS1 | PYM / WIBG |
| | NUP160 | UPF3A |
| | NXF5 | hNAG |
| | EIF5A | |
| | DDX25 | |
| | NUP133 | |
| | NUP107 | |
| *p*-value in GSEA | 0.012 | 0.0026 |

* Core enrichment genes

# Genes associated with NMD

## Supplementary Table 7. List of significantly unspliced exonic regions

| chr | start | end | symbol | #exonic reads (wt) | #junction reads (wt) | #exonic reads (mt) | #junction reads (mt) | odds ratio | q-value |
|---|---|---|---|---|---|---|---|---|---|
| chr1 | 153507075 | 153507306 | S100A6 | 33709 | 117 | 21629 | 478 | 6.37 | 1.70E-88 |
| chr16 | 70287170 | 70287284 | AARS | 907 | 52 | 1007 | 584 | 10.12 | 1.17E-65 |
| chr19 | 11561460 | 11561782 | PRKCSH | 3513 | 462 | 2704 | 894 | 2.51 | 1.26E-46 |
| chr12 | 57906018 | 57906136 | MARS | 1325 | 66 | 539 | 149 | 5.55 | 1.07E-28 |
| chr19 | 47278139 | 47279004 | SLC1A5 | 27539 | 22 | 23528 | 169 | 8.99 | 7.53E-28 |
| chr19 | 47280231 | 47280366 | SLC1A5 | 4340 | 14 | 3381 | 128 | 11.74 | 2.49E-24 |
| chr16 | 85840343 | 85840607 | COX4I1 | 9411 | 22 | 7463 | 144 | 8.25 | 3.49E-24 |
| chr11 | 65657874 | 65659106 | CCDC85B | 10482 | 210 | 6391 | 318 | 2.48 | 1.51E-21 |
| chr11 | 63970921 | 63971094 | STIP1 | 3925 | 25 | 4647 | 186 | 6.28 | 5.70E-19 |
| chr19 | 11560080 | 11560243 | PRKCSH | 2949 | 198 | 1832 | 307 | 2.50 | 5.70E-19 |
| chr11 | 62480096 | 62482851 | HNRNPUL2-BSCL2 | 7946 | 13 | 5596 | 90 | 9.83 | 1.42E-17 |
| chr4 | 48853822 | 48854037 | OCIAD1 | 1184 | 7 | 798 | 75 | 15.90 | 1.35E-16 |
| chr16 | 29819890 | 29820062 | MAZ | 4957 | 100 | 3891 | 224 | 2.85 | 1.28E-15 |
| chrX | 48340984 | 48341182 | FTSJ1 | 1604 | 242 | 1223 | 405 | 2.19 | 1.91E-15 |
| chr11 | 65338861 | 65339239 | SSSCA1 | 2378 | 28 | 1851 | 115 | 5.28 | 1.32E-14 |
| chr8 | 97244001 | 97244168 | UQCRB | 3609 | 153 | 3449 | 340 | 2.33 | 1.70E-14 |
| chr15 | 72494794 | 72494961 | PKM2 | 17689 | 214 | 9797 | 255 | 2.15 | 1.36E-13 |
| chr1 | 108681679 | 108681830 | SLC25A24 | 441 | 8 | 549 | 121 | 12.15 | 1.48E-13 |
| chr17 | 16284366 | 16284498 | UBB | 2952 | 57 | 1697 | 119 | 3.63 | 1.48E-13 |
| chr8 | 103848483 | 103848618 | AZIN1 | 324 | 18 | 319 | 127 | 7.17 | 1.70E-13 |
| chr12 | 7085346 | 7086053 | LPCAT3 | 1112 | 45 | 945 | 151 | 3.95 | 1.70E-13 |
| chr22 | 39712710 | 39712901 | RNU86;RPL3 | 7090 | 164 | 5374 | 280 | 2.25 | 1.70E-13 |
| chr11 | 8706264 | 8706439 | RPL27A | 224 | 34 | 162 | 134 | 5.45 | 4.70E-13 |
| chr12 | 120934204 | 120934356 | DYNLL1 | 2580 | 35 | 2335 | 135 | 4.26 | 5.08E-13 |
| chr16 | 30081437 | 30081735 | ALDOA | 29279 | 66 | 26261 | 179 | 3.02 | 1.31E-12 |
| chr15 | 43700140 | 43700286 | TP53BP1 | 315 | 75 | 410 | 313 | 3.21 | 1.56E-12 |
| chr12 | 120933858 | 120934019 | DYNLL1 | 937 | 20 | 764 | 97 | 5.95 | 2.78E-12 |
| chr11 | 43363977 | 43366080 | API5 | 5846 | 43 | 5900 | 157 | 3.62 | 3.45E-12 |
| chr12 | 57906533 | 57906747 | MARS | 3456 | 37 | 1671 | 78 | 4.36 | 3.46E-12 |
| chr2 | 9547577 | 9547727 | ITGB1BP1 | 1408 | 96 | 931 | 176 | 2.77 | 5.23E-12 |
| chr2 | 101619163 | 101619270 | RPL31 | 12724 | 60 | 9511 | 142 | 3.17 | 5.23E-12 |
| chr21 | 46396594 | 46396888 | C21orf70 | 868 | 71 | 772 | 191 | 3.02 | 1.12E-11 |
| chr15 | 43701094 | 43701294 | TP53BP1 | 820 | 79 | 1131 | 299 | 2.74 | 1.51E-11 |
| chr2 | 36775667 | 36778278 | CRIM1 | 10865 | 5 | 12951 | 84 | 14.09 | 4.88E-11 |
| chr9 | 90342941 | 90343064 | CTSL1 | 569 | 22 | 475 | 99 | 5.39 | 4.88E-11 |
| chr12 | 7084857 | 7085165 | EMG1 | 3198 | 50 | 2224 | 118 | 3.39 | 4.92E-11 |
| chr3 | 52561844 | 52561947 | NT5DC2 | 581 | 38 | 245 | 73 | 4.56 | 5.47E-11 |
| chr7 | 6628247 | 6628610 | ZDHHC4 | 1727 | 111 | 1698 | 260 | 2.38 | 7.01E-11 |
| chr1 | 206764973 | 206765177 | EIF2D | 470 | 13 | 392 | 81 | 7.47 | 7.56E-11 |
| chr9 | 130213559 | 130213711 | RPL12 | 13018 | 140 | 12089 | 279 | 2.15 | 7.60E-11 |
| chr3 | 49049067 | 49051549 | WDR6 | 19909 | 293 | 18843 | 481 | 1.73 | 9.95E-11 |
| chr7 | 107580433 | 107580803 | LAMB1 | 5277 | 29 | 2373 | 60 | 4.60 | 2.03E-10 |
| chr16 | 15129872 | 15131552 | PDXDC1 | 6455 | 15 | 7962 | 109 | 5.89 | 3.15E-10 |
| chrX | 153627827 | 153627935 | RPL10 | 11923 | 359 | 11323 | 559 | 1.64 | 4.40E-10 |
| chr14 | 20923289 | 20923553 | APEX1 | 2020 | 75 | 1831 | 182 | 2.68 | 4.87E-10 |
| chr5 | 176910394 | 176910731 | PDLIM7 | 3809 | 8 | 2438 | 49 | 9.57 | 5.04E-10 |
| chr1 | 153507677 | 153507836 | S100A6 | 38122 | 298 | 20244 | 286 | 1.81 | 5.17E-10 |
| chr12 | 7086295 | 7086424 | LPCAT3 | 159 | 23 | 148 | 120 | 5.61 | 6.10E-10 |
| chr12 | 48541571 | 48543781 | ASB8 | 2392 | 5 | 3224 | 86 | 12.76 | 9.99E-10 |
| chr19 | 15226376 | 15226482 | ILVBL | 158 | 13 | 107 | 72 | 8.18 | 9.99E-10 |
| chr1 | 155287959 | 155288082 | FDPS | 3023 | 102 | 2374 | 191 | 2.38 | 1.05E-09 |
| chr3 | 50324499 | 50325702 | IFRD2;C3orf45 | 4618 | 115 | 4415 | 244 | 2.22 | 1.09E-09 |
| chr12 | 120638532 | 120638694 | RPLP0 | 3728 | 61 | 1893 | 95 | 3.07 | 1.39E-09 |
| chr20 | 33516235 | 33516754 | GSS | 2679 | 45 | 2231 | 122 | 3.26 | 1.59E-09 |
| chr11 | 64084920 | 64085033 | TRMT112 | 1275 | 218 | 959 | 319 | 1.95 | 4.87E-09 |
| chr22 | 45588122 | 45593097 | KIAA0930 | 11755 | 5 | 15716 | 80 | 11.97 | 5.07E-09 |
| chr2 | 9545814 | 9547034 | ITGB1BP1 | 1862 | 72 | 1697 | 171 | 2.61 | 5.90E-09 |
| chr7 | 91736606 | 91736736 | AKAP9 | 107 | 7 | 167 | 117 | 10.71 | 6.56E-09 |
| chr17 | 4863054 | 4863211 | SPAG7 | 548 | 33 | 396 | 94 | 3.94 | 7.26E-09 |
| chr16 | 70286296 | 70286809 | AARS | 8954 | 490 | 10077 | 818 | 1.48 | 1.28E-08 |
| chr10 | 75530712 | 75531933 | SEC24C | 5885 | 5 | 5494 | 55 | 11.78 | 1.44E-08 |
| chr7 | 150933493 | 150933676 | CHPF2 | 269 | 21 | 229 | 90 | 5.03 | 1.53E-08 |
| chrX | 24094838 | 24096927 | EIF2S3 | 30820 | 49 | 25372 | 118 | 2.93 | 2.63E-08 |
| chr9 | 134397433 | 134397611 | POMT1 | 478 | 9 | 509 | 74 | 7.72 | 2.74E-08 |
| chr9 | 139757232 | 139757451 | EDF1 | 2055 | 47 | 1757 | 121 | 3.01 | 2.98E-08 |
| chr1 | 24120916 | 24121090 | LYPLA2 | 1029 | 70 | 913 | 162 | 2.61 | 3.37E-08 |
| chrX | 153628804 | 153628967 | RPL10 | 12270 | 592 | 8688 | 617 | 1.47 | 3.47E-08 |
| chr2 | 113414698 | 113414818 | SLC20A1 | 510 | 13 | 261 | 44 | 6.61 | 3.82E-08 |
| chr9 | 140082867 | 140083230 | ANAPC2;SSNA1 | 376 | 190 | 294 | 324 | 2.18 | 3.92E-08 |
| chr1 | 110563352 | 110563473 | AHCYL1 | 310 | 8 | 147 | 36 | 9.49 | 4.05E-08 |
| chr4 | 2249159 | 2252428 | MIR4800;MXD4 | 5251 | 12 | 3179 | 46 | 6.33 | 4.81E-08 |
| chr12 | 7086759 | 7086907 | LPCAT3 | 252 | 58 | 168 | 127 | 3.28 | 5.05E-08 |
| chr10 | 127524672 | 127525424 | BCCIP;DHX32 | 1934 | 30 | 1785 | 100 | 3.61 | 5.33E-08 |
| chr17 | 2267352 | 2267477 | SGSM2 | 120 | 3 | 86 | 44 | 20.47 | 8.09E-08 |
| chr1 | 45480110 | 45480248 | UROD | 1062 | 43 | 854 | 108 | 3.12 | 9.56E-08 |
| chr12 | 7087502 | 7087669 | LPCAT3 | 287 | 59 | 262 | 160 | 2.97 | 1.08E-07 |
| chr17 | 73126847 | 73126962 | NT5C | 350 | 63 | 294 | 151 | 2.85 | 1.51E-07 |
| chr13 | 47351590 | 47351758 | ESD | 1031 | 12 | 766 | 54 | 6.06 | 1.68E-07 |
| chr11 | 63971525 | 63972020 | STIP1 | 8481 | 13 | 10376 | 85 | 5.34 | 1.70E-07 |
| chr19 | 10506110 | 10506219 | CDC37 | 947 | 128 | 747 | 215 | 2.13 | 1.77E-07 |
| chr2 | 27535860 | 27535976 | MPV17 | 211 | 32 | 90 | 62 | 4.54 | 1.93E-07 |
| chr15 | 43699411 | 43699768 | TP53BP1 | 945 | 57 | 1466 | 225 | 2.54 | 2.34E-07 |
| chr22 | 40803204 | 40803332 | SGSM3 | 330 | 38 | 207 | 86 | 3.61 | 2.56E-07 |
| chr12 | 7088632 | 7088741 | LPCAT3 | 132 | 79 | 96 | 184 | 3.20 | 2.62E-07 |
| chr17 | 2266758 | 2266893 | SGSM2 | 163 | 3 | 89 | 31 | 18.93 | 3.00E-07 |
| chr2 | 233431787 | 233431924 | EIF4E2 | 832 | 29 | 868 | 106 | 3.50 | 3.86E-07 |
| chr5 | 6656190 | 6656292 | SRD5A1 | 366 | 4 | 262 | 38 | 13.27 | 3.86E-07 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| chr2 | 36774126 | 36774314 | CRIM1 | 830 | 13 | 946 | 79 | 5.33 | 4.45E-07 |
| chr6 | 90036343 | 90039676 | UBE2J1 | 3571 | 2 | 2998 | 36 | 21.44 | 6.89E-07 |
| chr22 | 45607836 | 45607988 | KIAA0930 | 719 | 189 | 891 | 429 | 1.83 | 6.92E-07 |
| chr5 | 172386867 | 172387044 | RPL26L1 | 1441 | 82 | 1151 | 152 | 2.32 | 7.44E-07 |
| chr20 | 44520207 | 44520401 | CTSA | 1232 | 62 | 1050 | 133 | 2.52 | 1.24E-06 |
| chr12 | 7086528 | 7086687 | LPCAT3 | 238 | 79 | 187 | 166 | 2.67 | 1.42E-06 |
| chr12 | 7091836 | 7091943 | LPCAT3 | 115 | 13 | 58 | 46 | 7.02 | 1.44E-06 |
| chr1 | 44463526 | 44463697 | SLC6A9 | 518 | 134 | 338 | 190 | 2.17 | 1.61E-06 |
| chr1 | 113454469 | 113456787 | SLC16A1 | 13506 | 97 | 14856 | 217 | 2.03 | 1.61E-06 |
| chr2 | 85570761 | 85570921 | RETSAT | 293 | 5 | 203 | 37 | 10.68 | 1.61E-06 |
| chr3 | 49060308 | 49060409 | NDUFAF3 | 436 | 17 | 212 | 42 | 5.08 | 1.64E-06 |
| chr6 | 13632601 | 13632753 | RANBP9 | 333 | 19 | 424 | 102 | 4.22 | 1.64E-06 |
| chr20 | 44523632 | 44523772 | CTSA | 2602 | 40 | 2362 | 105 | 2.89 | 1.64E-06 |
| chr16 | 89627634 | 89627776 | RPL13 | 7682 | 78 | 6689 | 152 | 2.24 | 1.66E-06 |
| chr16 | 85839338 | 85839470 | COX4I1 | 2157 | 33 | 1739 | 85 | 3.19 | 1.67E-06 |
| chr3 | 48646984 | 48647098 | UQCRC1 | 254 | 6 | 108 | 26 | 10.19 | 1.70E-06 |
| chr19 | 13035452 | 13035621 | FARSA | 1453 | 4 | 1184 | 38 | 11.66 | 1.70E-06 |
| chr11 | 119179233 | 119180625 | MCAM | 14236 | 129 | 9009 | 162 | 1.98 | 1.72E-06 |
| chr1 | 45479265 | 45479463 | UROD | 1876 | 53 | 1437 | 107 | 2.64 | 2.11E-06 |
| chr7 | 33053741 | 33054443 | NT5C3 | 1128 | 4 | 1675 | 66 | 11.11 | 2.11E-06 |
| chr7 | 91737807 | 91737947 | AKAP9 | 163 | 24 | 317 | 176 | 3.77 | 2.11E-06 |
| chr10 | 75530037 | 75530229 | SEC24C | 616 | 23 | 553 | 79 | 3.83 | 2.11E-06 |
| chr5 | 466841 | 467409 | EXOC3 | 4138 | 24 | 3191 | 68 | 3.67 | 2.16E-06 |
| chr19 | 10943113 | 10943889 | TMED1 | 3099 | 84 | 3189 | 185 | 2.14 | 2.58E-06 |
| chr9 | 139757739 | 139757900 | EDF1 | 6635 | 3 | 4832 | 31 | 14.19 | 2.94E-06 |
| chr5 | 917122 | 918164 | TRIP13 | 8558 | 32 | 11737 | 130 | 2.96 | 3.32E-06 |
| chrX | 40988252 | 40988398 | USP9X | 184 | 0 | 232 | 45 | NA | 3.33E-06 |
| chr14 | 102348497 | 102348608 | PPP2R5C | 307 | 1 | 282 | 35 | 38.10 | 3.43E-06 |
| chr11 | 64084735 | 64084837 | TRMT112 | 2159 | 338 | 2117 | 512 | 1.54 | 4.17E-06 |
| chr1 | 3548029 | 3548221 | WRAP73 | 236 | 61 | 159 | 118 | 2.87 | 4.47E-06 |
| chr6 | 33646197 | 33646320 | ITPR3 | 481 | 12 | 515 | 67 | 5.21 | 4.91E-06 |
| chr2 | 99226218 | 99226502 | UNC50 | 785 | 16 | 570 | 53 | 4.56 | 5.05E-06 |
| chr4 | 48862741 | 48863834 | OCIAD1 | 3130 | 3 | 2262 | 30 | 13.84 | 5.05E-06 |
| chr15 | 72492814 | 72492996 | PKM2 | 29133 | 262 | 17779 | 262 | 1.64 | 5.05E-06 |
| chr12 | 111884556 | 111884658 | SH2B3 | 39 | 3 | 25 | 41 | 21.32 | 5.18E-06 |
| chr6 | 31929685 | 31929831 | SKIV2L | 303 | 3 | 176 | 26 | 14.92 | 5.57E-06 |
| chr11 | 64082472 | 64082742 | ESRRA | 843 | 13 | 656 | 51 | 5.04 | 5.57E-06 |
| chr11 | 71941405 | 71941512 | INPPL1 | 1190 | 73 | 833 | 120 | 2.35 | 6.32E-06 |
| chr3 | 50326877 | 50327059 | IFRD2 | 1446 | 71 | 952 | 111 | 2.37 | 6.44E-06 |
| chr6 | 31605222 | 31605554 | PRRC2A | 675 | 22 | 629 | 76 | 3.71 | 7.50E-06 |
| chr10 | 75542080 | 75542199 | CHCHD1 | 815 | 31 | 771 | 92 | 3.14 | 7.50E-06 |
| chr12 | 7090165 | 7090344 | LPCAT3 | 303 | 11 | 201 | 43 | 5.89 | 7.50E-06 |
| chr4 | 3521798 | 3521920 | LRPAP1 | 475 | 0 | 220 | 15 | NA | 8.84E-06 |
| chr7 | 33055281 | 33055482 | NT5C3 | 165 | 10 | 211 | 76 | 5.94 | 8.90E-06 |

For each exonic region, reads mapped on the exonic region and on surrounding junction regions are counted for both wild type and mutant cells experiments.
The significance for the difference of the exonic and junction reads ratios between wild type and mutatnt cells is measured by the p-value of Z test.
Then, false discovery rate (FDR) q-values are estimated using ***qvalue*** R package (version 1.26.0) [http://cran.r-project.org/web/packages/qvalue/index.html].
Exonic regions with q-values below 1.0E-5 are listed above.

**Supplementary Table 8. List of primers used for validation of somatic mutations detected in whole exome analysis**

| Gene | Forward Primer | Reverse Primer | Sample Name |
|---|---|---|---|
| *DAPK1* | 5'-TTGGTTAGTGGTGTTTCACGA-3' | 5'-GATTCCTGGCCAACCACA-3' | MDS-03 |
| *LRP1* | 5'-TTCTGCCGAACTGTCCTTC-3' | 5'-TTCTTGGCCTTGCAGGAG-3' | MDS-03 |
| *TBX19* | 5'-ATGGCCATGAGTGAGCTG-3' | 5'-GCGGCAGCCTCTTCAATA-3' | MDS-03 |
| *TTF2* | 5'-TCCCTTGTTTTCTGAGCAGG-3' | 5'-TCACAGGGGTGGCTATTCA-3' | MDS-03 |
| *U2AF35* | 5'-GCCTCGTGTGCATTCTCTG-3' | 5'-CTTTTCAGTTTCGCCGTGAG-3' | MDS-03 |
| *DDX4* | 5'-GCACAGCCTCTAGTAAAAGTATTGA-3' | 5'-GGCACGAAAACAATGGCT-3' | MDS-04 |
| *LAMA1* | 5'-CCTAAAATGCAGGCAGCG-3' | 5'-CTGCTGCACTGGGTGAAA-3' | MDS-04 |
| *LRRC46* | 5'-TCCTTCCCACAAGGGACA-3' | 5'-CCCTTCCTTCCCCAGAAA-3' | MDS-04 |
| *MC5R* | 5'-ATCATCGCCGGCATCT-3' | 5'-AGCAGCATGGTGACGGT-3' | MDS-04 |
| *C11orf61* | 5'-TCTGATCTTCCCAATGGTCA-3' | 5'-AAAAGCCTGGTTTTGGAAAG-3' | MDS-06 |
| *C6orf170* | 5'-AAAGTCCCACTGAGTATCCACTAAT-3' | 5'-GAGAATTACTCTCCTGCAAGTATGG-3' | MDS-06 |
| *COL4A5* | 5'-GAAGGCTTCCAATGAAGCAG-3' | 5'-TGACAAATGCAAGGAAGAGTG-3' | MDS-06 |
| *HIVEP2* | 5'-ACAATGGCAGGGCTATTCTG-3' | 5'-ATTCAGGAGCATCAGGGCTA-3' | MDS-06 |
| *KCNQ3* | 5'-GGGCTGACTTTGTCAATGGT-3' | 5'-TATGCACATGCAACACATGG-3' | MDS-06 |
| *LRRFIP1* | 5'-CTGTGGGTTGGGGTCAGTT-3' | 5'-TAAAGGGCTTGGACCAGAAA-3' | MDS-06 |
| *TET2* | 5'-TGCTGATGATGCTGATAATGC-3' | 5'-TCATTGTCCCTGCAGTCTGT-3' | MDS-06 |
| *ZRSR2* | 5'-GACGTTTGGAATGGAGCAGT-3' | 5'-TGTGTCCCAGCTCTCTTGTG-3' | MDS-06 |
| *ARMC9* | 5'-TGCTGAATTTGAAGACACCTTG-3' | 5'-TGCCATCTAAAGTAACTCTGAGCA-3' | MDS-07 |
| *ATP6V0A2* | 5'-GCAGTAACCCAGATTCGTTGA-3' | 5'-TGCCATTGGAAATACGGG-3' | MDS-07 |
| *CDH11* | 5'-AGAGAGCTGGCTGTGTCACC-3' | 5'-TCATCAGAACAGCCCTACCC-3' | MDS-07 |
| *CPAMD8* | 5'-ACAGGAACCTTCTCCCTGCT-3' | 5'-TGCGTGTCTTCCATAGGTTG-3' | MDS-07 |
| *F5* | 5'-CACTTGTCCATATTCGCTGG-3' | 5'-TGCCTGAGCATTACTTTCACTC-3' | MDS-07 |
| *KBTBD5* | 5'-ACTTGGATTGCAAAGGGTTG-3' | 5'-CTCTGTGGGAACCAGCTCTC-3' | MDS-07 |
| *KRT37* | 5'-GAATTGCCCAGATCACACG-3' | 5'-TGAGTTTTGGGGCACCTG-3' | MDS-07 |
| *MCRS1* | 5'-AGCAGCGGAGCCTAAGAAG-3' | 5'-GCCAAATTAGAGCCCAGC-3' | MDS-07 |
| *NONO* | 5'-TGGTCCTCACAAGGAATGC-3' | 5'-CTCATGGCGTGCAGCTT-3' | MDS-07 |
| *PDZRN3* | 5'-CGGATCTTCACCTTCCACTC-3' | 5'-CTGCCCTCCTATCACCACTC-3' | MDS-07 |
| *RFX6* | 5'-CCCATGGGAAACATTGCTTA-3' | 5'-GTCTCCACGTGGCTTTCATT-3' | MDS-07 |
| *SF3A1* | 5'-ACAAGGTCCTGCCTCAGAGA-3' | 5'-TGCCTCTTTGTACTGCGTTG-3' | MDS-07 |
| *TNKS1BP* | 5'-CTCTGTCGTTGCCAATGATG-3' | 5'-GTGATGCTAGCCAGGAGGTG-3' | MDS-07 |
| *TP53BP1* | 5'-GCTGCAACTCCTGGTCAAGT-3' | 5'-AGGGAGTGGAGAAGTGCTCA-3' | MDS-07 |
| *C1QTNF3* | 5'-GGTCAGGCAGGTTGGAAA-3' | 5'-GACTGGGTTTTCTCATGTGATG-3' | MDS-08 |
| *CASR* | 5'-GCCTGTAACAAGTGCCCAG-3' | 5'-TGGAGAAGCAGCAGAGCA-3' | MDS-08 |
| *CKLF* | 5'-CACAAGCCCCTGAACCAT-3' | 5'-GAGCTCCTCCTTCCTAAGCC-3' | MDS-08 |
| *CRKL* | 5'-CCCACCAATGGGATCTGT-3' | 5'-TGCTTTCCGTGTGGTGAG-3' | MDS-08 |
| *DOT1L* | 5'-AAGCCGTCTCCCTCCAAA-3' | 5'-CAGGGACACCCTGAACACA-3' | MDS-08 |
| *ETNK1* | 5'-TCGGGAGACTGACTAAAGTTTCA-3' | 5'-CCCTATCATAGAGCAAAGCAATG-3' | MDS-08 |
| *LRSAM1* | 5'-TCCTTTGCAGCTGCTTACA-3' | 5'-TAAAGCCCCAGTCCACCTC-3' | MDS-08 |
| *TEP1* | 5'-TGTTAACCCTGCCCACTGTC-3' | 5'-TTCTCTAGGTTCCCAGGGGG-3' | MDS-08 |
| *TNK2* | 5'-TGAGGACAGAGGACGGAAA-3' | 5'-ACAACCCTTGTCCCTGAGC-3' | MDS-08 |
| *ZRSR2* | 5'-GAAATGTACCTTCGGAAAAGGA-3' | 5'-CCCTCCTTTCGGAGTTCTTC-3' | MDS-08 |
| *C11orf9* | 5'-TGGCCCAAACAGGCTAGA-3' | 5'-TGAGAGGCAAGGAACCAGG-3' | MDS-09 |
| *CALCR* | 5'-AATGAAATTTGTACAGAGGCCC-3' | 5'-TGCAGCACTGACTATGAATGG-3' | MDS-09 |
| *FAM47C* | 5'-AGACTCGGGTGTCCTGTCTC-3' | 5'-GAGGCTCCAGGTAGAGATGG-3' | MDS-09 |
| *GRID1* | 5'-CTGGTCCCAAGGTTTCCA-3' | 5'-GCGCTACAAAGGGTTCTCC-3' | MDS-09 |
| *MANBA* | 5'-AGCTTTGCGGGACTGCTA-3' | 5'-TGCCTATTTCATGTGGAAGACT-3' | MDS-09 |
| *SLCO4C1* | 5'-GGAAGAGATGTAAACCCTTACCA-3' | 5'-GCCTGATTTCATCAAGCTACG-3' | MDS-09 |
| *SFRS2* | 5'-AGACGCCATTTCCCCAG-3' | 5'-TTCGAGAAGTACGGGCG-3' | MDS-09 |
| *TP53* | 5'-AAATCGGTAAGAGGTGGGC-3' | 5'-AAAAAGGCCTCCCCTGC-3' | MDS-09 |
| *ZKSCAN1* | 5'-CGGCTGAAGGAACTTTGTC-3' | 5'-TTCAACGTCTGTGCCCTG-3' | MDS-09 |
| *EIF2AK2* | 5'-CGGAGTAAGCCAAACTCCTC-3' | 5'-TCTTTAAGGTTTGGCATGGA-3' | MDS-10 |
| *ARIH2* | 5'-TGGAGGGTGTTTGTGGCT-3' | 5'-GGGGTCATGAAAGGCAGA-3' | MDS-11 |
| *FAM175A* | 5'-CGTGACAATCTGATGCGAC-3' | 5'-TTGCAAGATAGCAAATGCC-3' | MDS-11 |
| *FBXL20* | 5'-GCCATTTTTACCCTGAGTCTCTT-3' | 5'-CCTTGCTCTATTTGGCTGC-3' | MDS-11 |
| *FER1L6* | 5'-GTGGACACCTAAATGTTTTCTTCC-3' | 5'-AACTGCCACAGGTAGGGGA-3' | MDS-11 |
| *LARP1B* | 5'-GAACGTCTTTTTCACTCCATTG-3' | 5'-AGCCTCTGGCTTTGAGCA-3' | MDS-11 |
| *PRAMEF11* | 5'-CACCATTTTCAGGATGCTTCT-3' | 5'-GCTGCTATGCAGGATGTGG-3' | MDS-11 |
| *RFC1* | 5'-TCTCAATGAGCTTCCCCAA-3' | 5'-TGTTGAGTGGGAAAATGTCA-3' | MDS-11 |
| *SLC20A1* | 5'-AACCAAATAAACTCCAGTGGC-3' | 5'-GCCATGTCAGCTTCTCCA-3' | MDS-11 |
| *STAC3* | 5'-TGGGAAGCCTAGGAGTTAAAGA-3' | 5'-TTGGATAGACGAGGCCCTT-3' | MDS-11 |
| *STAG1* | 5'-TGCTCCATGCTTCCTCAAC-3' | 5'-TGGAAAGGAGGGAGGATGT-3' | MDS-11 |
| *TTN* | 5'-AATGAGGCACCAGCCTTG-3' | 5'-AAGCCGCTGTTTGATGGT-3' | MDS-11 |
| *ATRX* | 5'-CCAACTTTGTTTCCCTCTCTG-3' | 5'-TCTCTAGGAATTAATCTGGTAGCTG-3' | MDS-12 |
| *C3orf63* | 5'-CCTGAACCAAAACCTCGC-3' | 5'-GGATCCTGGACTTTACAAGCA-3' | MDS-12 |
| *DACH1* | 5'-GACTCACCTTGCGTTGGTG-3' | 5'-GGCCAAAGTGGCTTCCTT-3' | MDS-12 |

**Continued from previous page**

| | | | |
|---|---|---|---|
| *DRD1* | 5'-AGGGGCTTGTGGGTTTTG-3' | 5'-TTCTGGCCCTTTGGGTC-3' | MDS-12 |
| *DYNC1H1* | 5'-ACGCACCTGCACAGATCA-3' | 5'-ACAGCCCGAAACACAACG-3' | MDS-12 |
| *EGF* | 5'-AAGGCTGTCCAACTATAAAACCC-3' | 5'-GAATATCCCCGAGCCAAG-3' | MDS-12 |
| *ENC1* | 5'-TTCAGGTGAGTGAGGACGC-3' | 5'-AGGTGGCATGGTCAACAA-3' | MDS-12 |
| *FAT2* | 5'-TGAGTAGGTGACATCTGCGTT-3' | 5'-GCAGAAACTGGATCGGGA-3' | MDS-12 |
| *FSIP2* | 5'-TGCTCAAGGACATATCTTCCG-3' | 5'-GGTGCTTCTTTGACTATGTGTTCA-3' | MDS-12 |
| *KBTBD4* | 5'-CAGCCACGATGAACACCA-3' | 5'-CCTCTCCAACGCAGTCATT-3' | MDS-12 |
| *KRT72* | 5'-TCCCTCCCAATTCTTGGTC-3' | 5'-AAGCACAGGGTGAAGGGA-3' | MDS-12 |
| *LPHN3* | 5'-CTCATGATTATGGCCAAGAGAA-3' | 5'-TGCTGCCCAGTTTCTTCA-3' | MDS-12 |
| *PLCXD3* | 5'-TAGCTTTGGGGGTCAGCA-3' | 5'-CGATTTTTGCCCAGGAAG-3' | MDS-12 |
| *ROCK2* | 5'-AGGACAACAGTGACACTCAACA-3' | 5'-GCGCTGCCTTACACAAAA-3' | MDS-12 |
| *SRL* | 5'-GACCATAGGCACATGCGA-3' | 5'-TTCCTTCACCTCTCACCCC-3' | MDS-12 |
| *U2AF35* | 5'-AGTCGATCACCTGCCTCACT-3' | 5'-GGTGCTTAATACCACGGAAAA-3' | MDS-12 |
| *UBE2Q1* | 5'-GGGCAGGTGCTGTGTTAAT-3' | 5'-GGCAGTAAGGCCACAGCTA-3' | MDS-12 |
| *ZNF518A* | 5'-TGGGCTTCAAGATGATGGA-3' | 5'-CCAATTCTGATGTTGCTGACA-3' | MDS-12 |
| *ZNF534* | 5'-GTCGAATTGCATTCCTTGC-3' | 5'-CCTTGCCACAGTCTATACATTCG-3' | MDS-12 |
| *CHRM2* | 5'-GGGAAGGGAGATTTGGGA-3' | 5'-CCATTTTTGTGGTCCGCT-3' | MDS-14 |
| *BRCC3* | 5'-GAGGTTGGGTGTCTTTTATTTG-3' | 5'-TGCAGCAAGCAGCAATG-3' | MDS-15 |
| *CHD7* | 5'-TTTGGCTCACTGCAACTCTG-3' | 5'-TGCACTCAATGACGTCCG-3' | MDS-15 |
| *COL24A1* | 5'-GGGGATTCAATGGTAGCAAA | 5'-GGGACAGCAATTATGCCCT-3' | MDS-15 |
| *KIAA0232* | 5'-CTGCAGGACCTTGGCAAT-3' | 5'-CCCAAGTTGAGTGTTTCAAAAG-3' | MDS-15 |
| *PCDHAC1* | 5'-TCACAAAAGTGGTAGCAGAGGA-3' | 5'-AACTGAGGGACAGAGTTGCTC-3' | MDS-15 |
| *SLC9A10* | 5'-GGCAGAACAGTGAGGGGAT-3' | 5'-TGCTGATAAGTGTGAGGGAGG-3' | MDS-15 |
| *SORBS2* | 5'-AAGAGGCCTCATGTTAGTTTGA-3' | 5'-GTCAAGAAAAGTACCATGCCA-3' | MDS-15 |
| *TET2* | 5'-TGCTGATGATGCTGATAATGC-3' | 5'-TCATTGTCCCTGCAGTCTGT-3' | MDS-15 |
| *U2AF35* | 5'-GCCTCGTGTGCATTCTCTG-3' | 5'-CTTTTCAGTTTCGCCGTGAG-3' | MDS-15 |
| *FYB* | 5'-CTCTTCCTGATTTATTTTGCTCTGG-3' | 5'-TCATCAAAAGGGTCCCCA-3' | MDS-16 |
| *LTBP2* | 5'-GCCTGATGTCACGGTGTCT-3' | 5'-GAATGAGTGTGATGACTTGAACG-3' | MDS-16 |
| *MPP2* | 5'-GTGAGCTGCTTGGTGGAT-3' | 5'-GGGCACCCTTTTCTCATC-3' | MDS-16 |
| *RABEP1* | 5'-CACCTGCCCTGAAGCTCTA-3' | 5'-TCACCTATTTCCTCCTTGCAG-3' | MDS-16 |
| *SRCAP* | 5'-TCCTCATGTACCCTTTCCAGA-3' | 5'-AGGGAACGGAGCAGCTCTT-3' | MDS-16 |
| *TET2_1* | 5'-TGCTGATGATGCTGATAATGC-3' | 5'-TCATTGTCCCTGCAGTCTGT-3' | MDS-16 |
| *TET2_2* | 5'-GTTTGGGATGGAATGGTGAT-3' | 5'-ACCAAAGATTGGGCTTTCCT-3' | MDS-16 |
| *ZRSR2* | 5'-CAGAATATGCAGTGATAAGTGCTG-3' | 5'-AACTTTGGCTCCGGCTG-3' | MDS-16 |
| *HSPH1* | 5'-CCAGCTGACTCCTCAATAAAATG-3' | 5'-GCCTCTTTTCTGTTTGTGTAGC-3' | MDS-17 |
| *APBA2* | 5'-CCTCTGCATGTCCCATCTTT-3' | 5'-ACCTTCAGATGCCCTGTGTC-3' | MDS-18 |
| *CKAP5* | 5'-TGCCATAAAACGTGTTCACC-3' | 5'-TGCCTTGGACAATCTCCTTT-3' | MDS-18 |
| *CLEC1A* | 5'-CAGACAACATGGGGACAGAG-3' | 5'-TCCAGTCTGACTTGGTCATTT-3' | MDS-18 |
| *HIVEP1* | 5'-CACGCACATACTATCAAACTGG-3' | 5'-TGGACTGCTTGGAGTGAAGC-3' | MDS-18 |
| *IDH2* | 5'-CTGCCTCTTTGTGGCCTAAG-3' | 5'-ATTCTGGTTGAAAGATGGCG-3' | MDS-18 |
| *NRP1* | 5'-CCATGCGGAGTTACAAGC-3' | 5'-CCAAAACAAAGCAGCCTCT-3' | MDS-18 |
| *NXF1* | 5'-CACACACTCCTGAGTTCTGACTG-3' | 5'-GCCCGGTTAGTATCACATCC-3' | MDS-18 |
| *PHF6* | 5'-GGGATGACAATATTAGAGGGCTTA-3' | 5'-TCTTGTACTCCACAGTGGTAATGG-3' | MDS-18 |
| *SFRS2* | 5'-AGACGCCATTTCCCCAG-3' | 5'-TTCGAGAAGTACGGGCG-3' | MDS-18 |
| *TBC1D9B* | 5'-GTGAACCTCTGGTGGCAGAT-3' | 5'-GTGACTCACCCCGGGTACTA-3' | MDS-18 |
| *TMEM146* | 5'-GGAGGAGGGAGGCTCAGG-3' | 5'-AGGCGTCCCCTGTCGTAG-3' | MDS-18 |
| *ADHFE1* | 5'-AAAATGATGCCCATGGCT-3' | 5'-TGAATACTTAGCATGTGCGTTG-3' | MDS-19 |
| *ANKRD34B* | 5'-TGCCTTGCAAGCACTAAGAA-3' | 5'-CCCCTTTAATGATCGCTTG-3' | MDS-19 |
| *ASXL1* | 5'-AGAGGACCTGCCTTCTCTGAGAAA-3' | 5'-TTCGATGGGATGGGTATCCAATGC-3' | MDS-19 |
| *BCOR* | 5'-GGAGGAGTCCAGGAAAAGG-3' | 5'-CCTCCCTGCTGCAGAAAT-3' | MDS-19 |
| *CSMD1* | 5'-TGCAGGCATATCCTCACG-3' | 5'-TCAATATGTTGCGCCTTCA-3' | MDS-19 |
| *DNMT3A* | 5'-TCTGTGGGGAAGGGAGA-3' | 5'-TTAGGCTCTGTGAGGCCAG-3' | MDS-19 |
| *EML1* | 5'-TCAAGTTCCTTCCTGCCTTT-3' | 5'-GGATACACCCATTCCAGCTT-3' | MDS-19 |
| *GTF3C4* | 5'-TTGGCGTTTTAAGCTTTTCC-3' | 5'-TGGGGAGCGAGTCATCA-3' | MDS-19 |
| *IRS4* | 5'-TCTGGCCACCACCTGAA-3' | 5'-CCCCAGGGCAAAGAAGAT-3' | MDS-19 |
| *KIAA1522* | 5'-CGGGCACTGTCAGTAAATCA-3' | 5'-CTCTATTAGGTTGTGTCTGCTGTATC-3' | MDS-19 |
| *LAMB4* | 5'-CCACACATTTGTCCAAGACAG-3' | 5'-CCTGGAAACATTTTGGTGAA-3' | MDS-19 |
| *NRAS* | 5'-CCGACAAGTGAGAGACAGGA-3' | 5'-GGCTCGCCAATTAACCCT-3' | MDS-19 |
| *PAPPA2* | 5'-TCCTTAGAGGCTGAAATGGAA-3' | 5'-CAGGGGTCATTGCAGGAT-3' | MDS-19 |
| *PVRL4* | 5'-TTTCTGGGTCATCTGCTGG-3' | 5'-TAGATGTTTTGCGGGGGA-3' | MDS-19 |
| *RUNX1* | 5'-TCTGCCAACTCCTTCATGC-3' | 5'-AAATCCCACCCCACTTTACA-3' | MDS-19 |
| *STAG2* | 5'-TCCTTTGCAGAGATGGCA-3' | 5'-TGCAATCCTACAATTCTGTGTG-3' | MDS-19 |
| *TET2_1* | 5'-ACTTTTCCTCACCCCCAAAG-3' | 5'-GCAGCATGCTTTTGAGTGTC-3' | MDS-19 |
| *TET2_2* | 5'-TGCTCAGCAAAGGTACTTGA-3' | 5'-AGTTCTGCAGCAGTGGTTTG-3' | MDS-19 |
| *ZNF225* | 5'-TCACATGCGAGAGAAACCA-3' | 5'-CTTGAAAGACCTGAGGCCC-3' | MDS-19 |
| *ALDH9A1* | 5'-TCTGGTTCAAAGGTACATGGG-3' | 5'-TGCAAGATCATGGAGATGTCA-3' | MDS-20 |

**Continued from previous page**

| | | | |
|---|---|---|---|
| *C10orf68* | 5'-GACAAATACGTGGAGATGGGA-3' | 5'-TGCTGCCAATAAAAAGGATG-3' | MDS-20 |
| *EDC4* | 5'-TGATGGAGCTGCTGCAAA-3' | 5'-TGTGGATGATGAAGCCCC-3' | MDS-20 |
| *MSLNL* | 5'-ATGGGTGGGACGGTCTA-3' | 5'-ATGGTGGCCTCCCTGTT-3' | MDS-20 |
| *NRAS* | 5'-CCGACAAGTGAGAGACAGGA-3' | 5'-GGCTCGCCAATTAACCCT-3' | MDS-20 |
| *NRIP1* | 5'-CTGGCACTAAACGTTGCAGA-3' | 5'-CCAGTGAGCACTCCACCTTT-3' | MDS-20 |
| *SEC14L4* | 5'-AAATGGTGACGTGGGACAA-3' | 5'-GACAACACCTACAGCCGGA-3' | MDS-20 |
| *ZNF583* | 5'-GCTTGAATGCAGGATAACCTATG-3' | 5'-CCTGGTGGAATGTTTGCC-3' | MDS-20 |
| *BTBD10* | 5'-CTGCAGTATTAGGCACAATTCA-3' | 5'-CCTCGCGTATTGCTAAAGG-3' | MDS-21 |
| *SF3B1* | 5'-TCCTCTGTGTTGGCGGAT-3' | 5'-TGCAGTTTGGCTGAATAGTTG-3' | MDS-21 |
| *ELANE* | 5'-AGCTTCCCCACCTTGTCTG-3' | 5'-CGGAGGGTTGGATGATAGAG-3' | tAML-01 |
| *EML4* | 5'-ACCTGTCCAGTTGCTCTGCT-3' | 5'-CTTCATGGCCACATAAACACA-3' | tAML-01 |
| *EZH2* | 5'-TTCTGTCAGGCTTGATCACC-3' | 5'-CTCGTTTCTGAACACTCGGC-3' | tAML-01 |
| *FBXO11* | 5'-AGCACTCCATGTCCATTGTC-3' | 5'-GCCATTACCTCCTTACTCGG-3' | tAML-01 |
| *FILIP1* | 5'-CCTCTGCCTTCTGCAGATTT-3' | 5'-GCTCAAAGCCATTACTTCCAA-3' | tAML-01 |
| *FLT3* | 5'-TGCAGAAACATTTGGCACA-3' | 5'-TCATTGTCGTTTTAACCCTGC-3' | tAML-01 |
| *GLE1* | 5'-CGGAAGGAAGAAGGCCA-3' | 5'-CCCCCAATAAGCAGATAGCA-3' | tAML-01 |
| *LAMA5* | 5'-ATGGAGAGGTGAAGGGTGGT-3' | 5'-CTCTGCGTGACAAGAAGGTG-3' | tAML-01 |
| *MAGEB1* | 5'-GGAGGCAGGAATGCTGAT-3' | 5'-TGCTCAGGTTTCCATCATTG-3' | tAML-01 |
| *OR8H1* | 5'-CGTAGCGATCATAGGCCATT-3' | 5'-GACCTCCAGCTTCACACTTCC-3' | tAML-01 |
| *PTX* | 5'-CAGCACCAGCTTGTTGTCAT-3' | 5'-GACCCTCCCTCCTCTGTTTC-3' | tAML-01 |
| *SH3CT2* | 5'-TGGCATTTGCTGGAACATTA-3' | 5'-TTCAGAAAGGGGGTCAAATG-3' | tAML-01 |
| *TET2_1* | 5'-CATGTATGCAGCCCTTCTCC-3' | 5'-ACATTTGTGACTTGTGCTCCA-3' | tAML-01 |
| *TET2_2* | 5'-TGCTCAGCAAAGGTACTTGA-3' | 5'-AGTTCTGCAGCAGTGGTTTG-3' | tAML-01 |
| *APOL6* | 5'-CACCGCTGGTCAAGGTTT-3' | 5'-CCAAAATGCACGGACGTT-3' | tAML-02 |
| *DENND2A* | 5'-TGGGACACAAACCACTGAGA-3' | 5'-ACGTTCAAACATGCCGGA-3' | tAML-02 |
| *ETV6* | 5'-AGACAATCAGTCAACCCAAGC-3' | 5'-ACAGTCGAGCCAGTCCGTT-3' | tAML-02 |
| *FAM13A* | 5'-CCAGCACTCAAAACTGAAAGA-3' | 5'-CAAGGGGACTTATCAACAAAGA-3' | tAML-02 |
| *FAM83B* | 5'-AACCATATCCGCTTTTTGC-3' | 5'-CTTTTGCTCAGGTAAAGTGGG-3' | tAML-02 |
| *HELLS* | 5'-GGCCCTTCCTCTTACTTTCA-3' | 5'-CCACAACAGCTCTACCAAAACA-3' | tAML-02 |
| *IMP4* | 5'-TCGGCAGACAGCGACAT-3' | 5'-CTGTGTGTCCTGGGTGTCTC-3' | tAML-02 |
| *KIAA1217* | 5'-GAATTACCAGTCATCTGCATTG-3' | 5'-GCCGTGCATGAGGAATAC-3' | tAML-02 |
| *KRT14* | 5'-ATCTCGCTCTTGCCGCT-3' | 5'-AATGTGGAGATGGACGCTG-3' | tAML-02 |
| *NCAPH* | 5'-GCAGCATTATTTACATGAGCC-3' | 5'-CCCCAAGGAAGCAATGAC-3' | tAML-02 |
| *PCDHB7* | 5'-CGGGGAGGGGAATATCTATC-3' | 5'-ATCTCTGGCTGACACGGAG-3' | tAML-02 |
| *PDS5B* | 5'-GGCCTTTGGAAGTGAGTTTG-3' | 5'-GATGCATCACCAAATGTGAA-3' | tAML-02 |
| *PHF8* | 5'-TTCCTCCTGACCACCCCTA-3' | 5'-TGGTACATTCAGAGACTCTTGGAG-3' | tAML-02 |
| *POLR2A* | 5'-TGCCACCCAGATGACCTT-3' | 5'-CATGAGCCACCATACCCAG-3' | tAML-02 |
| *RUNX1* | 5'-AAGGTTGAACCCAAGGAATCT-3' | 5'-GGCCACCAACCTCATTCT-3' | tAML-02 |
| *SAA4* | 5'-CAGAGGAGCAGGAAGCTCA-3' | 5'-AGGTGGAGCTGTCTCAGGTG-3' | tAML-02 |
| *SCML2* | 5'-CTTTGAGGAGAACACAAGGAAA-3' | 5'-TCTCATCTCTCTGAGGTTACTGACA-3' | tAML-02 |
| *ZMYND8* | 5'-CCTGCTCCTTATTGTTACTGTTC-3' | 5'-AATGCACTCTGTTCCCACA-3' | tAML-02 |
| *ANPEP* | 5'-CCTGTTTCCTTCTCCGCA-3' | 5'-CCCACTCCCAAAATCAGG-3' | tAML-03 |
| *GNB1* | 5'-CAGCAAATGGGTTCAGGTACT-3' | 5'-GAGGAGAATCGTTTAATCACAGATG-3' | tAML-03 |
| *ITGAD* | 5'-GAGTTCCTGGGATTCTGCC-3' | 5'-CAGCAGTTGTTACCTGGGC-3' | tAML-03 |
| *L1CAM* | 5'-TATTCTAGGGCCACGGCA-3' | 5'-AGTGCTCTCACTCGCACCT-3' | tAML-03 |
| *TP53* | 5'-TCAACTCTGTCTCCTTCCTCTTC-3' | 5'-AACCAGCCCTGTCGTCTCT-3' | tAML-03 |
| *ATM* | 5'-GCTTATCTGCTGCCGTCAA-3' | 5'-GCCAAATTCATATGCAAGGC-3' | tAML-04 |
| *ATP1A2* | 5'-AAGGGGATTCCCAAGCCT-3' | 5'-TACCTTGATGCCTGCGCT-3' | tAML-04 |
| *C1orf145* | 5'-AGTGTTTTTCACAAACCTCCTG-3' | 5'-GCGTCATTAGAAGGGGCTT-3' | tAML-04 |
| *CACNB2* | 5'-GCACCTAAATTGTGCATGCTT-3' | 5'-ATGGCAAAGGCTCTACGC-3' | tAML-04 |
| *KCTD1* | 5'-GGGTATTTGGTGAGGGTGG-3' | 5'-TGCCGGTTTTTCAGGTTT-3' | tAML-04 |
| *MKRN3* | 5'-TGAATACCCTGAGGGCTGG-3' | 5'-AGACCACATGCCACAGCA-3' | tAML-04 |
| *OR52H1* | 5'-AAACATGAGGATGACACAGACA-3' | 5'-CGCATCATACCCCACACA-3' | tAML-04 |
| *PRPF40B* | 5'-GGACTTGGGAACCCTGAGA-3' | 5'-GCAAAGACAAGGAAAGTCCAG-3' | tAML-04 |
| *AGAP1* | 5'-TGTGGTTCTGTCGTTGTGAAG-3' | 5'-CTTGATGAGCTCGACTGGC-3' | tAML-05 |
| *MTUS2* | 5'-TGTTGAGAAGGAAACCCAGG-3' | 5'-GCATTCGCAGCTACTGAGC-3' | tAML-05 |
| *NRAS* | 5'-CCGACAAGTGAGAGACAGGA-3' | 5'-GGCTCGCCAATTAACCCT-3' | tAML-05 |
| *OR1I1* | 5'-TTGCTTTTGGCATTGTCG-3' | 5'-GGATGAACACCCCACACA-3' | tAML-05 |
| *RAI2* | 5'-AGACAATGACCTTGGCCG-3' | 5'-ACCGGAAATCCGAGCCT-3' | tAML-05 |
| *SPATA20* | 5'-CTCAGTGACCTGCCCTTACC-3' | 5'-TGCCTGCACACACACAGA-3' | tAML-05 |
| *TBX5* | 5'-GGCACGCCATGAGAGTAGA-3' | 5'-GGACATCAGCTGCAACACG-3' | tAML-05 |
| *VPS13B* | 5'-CCTTGCTTCAGGGTCCTTC-3' | 5'-AGCACAACTGCAGAGTAAGCC-3' | tAML-05 |
| *ZRSR2* | 5'-CGCCTTAAATAGATGCAAGAGTC-3' | 5'-TCCAGTGGAAAAATCCCAGA-3' | tAML-05 |
| *PAMR1* | 5'-GCTGCTATATGCGAGGCTCT-3' | 5'-TGTTACTGACCTGGGGAAGG-3' | tAML-06 |
| *SPATS1* | 5'-TCATAAGCGAGGGCAATTCT-3' | 5'-TGGAGGAGGGTCTGTGACTT-3' | tAML-06 |
| *TRPM4* | 5'-CACATCTCCCCACAGATGAA-3' | 5'-CTCAGAGGAAGGCTGTCAGG-3' | tAML-06 |

**Continued from previous page**

| | | | |
|---|---|---|---|
| *KRAS* | 5'-TGTATCAAAGAATGGTCCTGC-3' | 5'-GGTACTGGTGGAGTATTTGATAGTG-3' | tAML-06 |
| *CACNA1E* | 5'-TTGCTTGATTCACTCTTTCCC-3' | 5'-TAGGCTGCCCCTTGAATG-3' | tAML-07 |
| *CNGA1* | 5'-CAAACGGCCAAATTCAGG-3' | 5'-CAAGGACTGCTGGTAAAGGAA-3' | tAML-07 |
| *DOCK4* | 5'-GCACTGCAGTTTGATGGC-3' | 5'-AAACACTTGCCACGGTCC-3' | tAML-07 |
| *GUCA1A* | 5'-CTTGGGTTATGATGGGCG-3' | 5'-AACCCTGGGCTCTCAGTTC-3' | tAML-07 |
| *HTR1A* | 5'-ATGATGCCCAGCGTCTTC-3' | 5'-TGAATGGAGAGTCGGGGA-3' | tAML-07 |
| *IDH1* | 5'-TGTGTTGAGATGGACGCCTA-3' | 5'-GGTGTACTCAGAGCCTTCGC-3' | tAML-07 |
| *IGSF3* | 5'-ACCTCGAGGGTTCCATCTG-3' | 5'-TCACTGTTCGGCTGGAGA-3' | tAML-07 |
| *LIX1* | 5'-TGGGTGAGGAAGTGTCAGG-3' | 5'-GGGTGCCTTTGTTGCTTTT-3' | tAML-07 |
| *MAST4* | 5'-CAGCTCAGATGAGTGCCGT-3' | 5'-GGAGAACTGGTGGCACAGA-3' | tAML-07 |
| *MED12* | 5'-CAGGGTGGGAGACACAAGA-3' | 5'-TAGCCCTGGAGGGAAGAAG-3' | tAML-07 |
| *OR51S1* | 5'-TCATTTCATGCTTGGTGAGC-3' | 5'-CCATCCTGAGCTGCCAAT-3' | tAML-07 |
| *SEZ6L* | 5'-TCCAAGCCTGAATCTGAGC-3' | 5'-CGCTGTAGTACAAAACTGCCA-3' | tAML-07 |
| *SSX7* | 5'-TGAGACGGGAATGATTTGG-3' | 5'-AATTGGCATTTGGCCTGT-3' | tAML-07 |
| *TET2_1* | 5'-TGCTGATGATGCTGATAATGC-3' | 5'-TCATTGTCCCTGCAGTCTGT-3' | tAML-07 |
| *TET2_2* | 5'-AATCCCATGAACCCTTACCC-3' | 5'-CCGGAGCTGCACTGTAGTTA-3' | tAML-07 |
| *TNS4* | 5'-TCCACCAACAAGGTAGGAGTG-3' | 5'-CGAAGTCTCCGTTGTGAGC-3' | tAML-07 |
| *GRIN2B* | 5'-TCACCGTCGTGGAGTAGTTG-3' | 5'-AGCACGTAGACCTGACCGA-3' | tAML-07 |
| *ZNF211* | 5'-GAGAGTTCATACTGGAGAAAGGC-3' | 5'-TCTCTGGTGGTAGATCAGGCT-3' | tAML-07 |
| *ZNF254* | 5'-CACAACTGCCCAGAGCAA-3' | 5'-AAGCTTCACCACATTCTTCACA-3' | tAML-07 |
| *ASXL1* | 5'-AGGTCAGATCACCCAGTCAGTT-3' | 5'-TAGCCCATCTGTGAGTCCAACTGT-3' | CMML-01 |
| *CBL* | 5'-AGGACCCAGACTAGATGCTTTC-3' | 5'-GGCCACCCCTTGTATCAGTA-3' | CMML-01 |
| *CSMD2* | 5'-AGCCACCCTTGAGGGAGAAG-3' | 5'-AACAGGGGAATGGTCGAAAG-3' | CMML-01 |
| *EZH2* | 5'-CCGTCTTCATGCTCACTGAC-3' | 5'-AAAAACCCTCCTTTGTCCAGA-3' | CMML-01 |
| *FMN2* | 5'-TTGAGTTGAGGTTTGCTCACTT-3' | 5'-TGGGCCTGATTAAAGGCA-3' | CMML-01 |
| *GNL2* | 5'-TGTGCAGAATGAAAAGCAGA-3' | 5'-TCTGTGAGTTTGTCGCAGGT-3' | CMML-01 |
| *ITPKC* | 5'-GGGTAAGCCCTCCAGTTTG-3' | 5'-CACCTTGATGCCCTCGAT-3' | CMML-01 |
| *KANK4* | 5'-CTGCCCAGCAACAGCTACTT-3' | 5'-TTTGCCAAGCAGAGAAGGTT-3' | CMML-01 |
| *KRAS* | 5'-TGTATCAAAGAATGGTCCTGC-3' | 5'-GGTACTGGTGGAGTATTTGATAGTG-3' | CMML-01 |
| *LUC7L2* | 5'-GCTCTCAGAAAATACTTTCCAGTTG-3' | 5'-CCACTTTTGTGAGGAAACCC-3' | CMML-01 |
| *MID2* | 5'-GCATCACCTGTGAGGTCTCC-3' | 5'-AATCGATCATTCAGGGATGC-3' | CMML-01 |
| *PCDHGC3* | 5'-AAATCAGTACTGCCCGTCCA-3' | 5'-GACCACAAACCCCACAGAAA-3' | CMML-01 |
| *TET2* | 5'-TGACCTCCAAACAATACACTGG-3' | 5'-CGAACTCGCTTGATTTTGAA-3' | CMML-01 |
| *U2AF35* | 5'-GCCTCGTGTGCATTCTCTG-3' | 5'-CTTTTCAGTTTCGCCGTGAG-3' | CMML-01 |
| *BCOR* | 5'-TTTTCCTGCCAGGTTTCTCT-3' | 5'-AGTTATTTCTCCATTAGATGACTGG-3' | CMML-02 |
| *C1orf101* | 5'-GCTGAATCTTGAGCACATGG-3' | 5'-GGATTTGCCACCAACCTG-3' | CMML-02 |
| *DNMT3A* | 5'-TCTCCATCCTCATGTTCTTGG-3' | 5'-TAAGCAGGCGTCAGAGGAGT-3' | CMML-02 |
| *ERBB4* | 5'-GGAGCTTCTGGAATTGTGCT-3' | 5'-TTTGTAACTCCATTAGCCACTGG-3' | CMML-02 |
| *HYDIN* | 5'-TTGCTGCCAGATGGGTTT-3' | 5'-CTTATTCGAGCCCAGCCA-3' | CMML-02 |
| *LRRC37A3* | 5'-GCAATTTGGACTCACGACAA-3' | 5'-TACCTTTTTCCCACCCCTGT-3' | CMML-02 |
| *LY75* | 5'-AACCAATAATGGGTGAAAGTTACTG-3' | 5'-GGAGGCTATGGGGAATTAACA-3' | CMML-02 |
| *MYST1* | 5'-TGTGCAGGATTCCTTTTGTT-3' | 5'-TTCTTCTGTACAGCATCCTTCA-3' | CMML-02 |
| *PANX3* | 5'-GGAAGAATTCAGCTGCTCCA-3' | 5'-TGTGCTTCTGGGTGGTTGTA-3' | CMML-02 |
| *SLIT3* | 5'-CAGCCTTTAAGTGGCACAGC-3' | 5'-TGGCAACTGGGACCTCATAC-3' | CMML-02 |
| *TRPC4* | 5'-TGCAAAATGCTAGCAACACAA-3' | 5'-TTTCCTTCTGGGTCTGACCTT-3' | CMML-02 |
| *U2AF35* | 5'-AGTCGATCACCTGCCTCACT-3' | 5'-GGTGCTTAATACCACGGAAAA-3' | CMML-02 |
| *UNC5D* | 5'-TGGCATTTGCTGGAACATTA-3' | 5'-TTCAGAAAGGGGGTCAAATG-3' | CMML-02 |
| *C14orf38* | 5'-CCCCATAATGATGGCATGT-3' | 5'-CACGGTGGGCTAAGCAAT-3' | CMML-03 |
| *CCDC83* | 5'-CAGAATTGCAACCCACAGA-3' | 5'-CAGCATTTCTCAAAACAGCA-3' | CMML-03 |
| *CYP7B1* | 5'-TGGCTGGAAAGAAGGGTTC-3' | 5'-GCAGTGCAATCCATGCAG-3' | CMML-03 |
| *EIF4G3* | 5'-AAACCCTGGTGGATTCCG-3' | 5'-CATGAGAAACTGCCATAATACAGC-3' | CMML-03 |
| *EPAS1* | 5'-AGAAGACAGAGCCCGAGCA-3' | 5'-TGGAGACATGGGGTGGA-3' | CMML-03 |
| *HDAC9* | 5'-TGTGCTCCTGAGAATTTGCTT-3' | 5'-CCCAAATGAATACTGGCCTC-3' | CMML-03 |
| *MYO1H* | 5'-AGACCAAAGGTTGAGAGAAGAA-3' | 5'-CAATCTACAGGACTGAAATTGGTAG-3' | CMML-03 |
| *SPRYD5* | 5'-CGGATTTTCACCAACCCA-3' | 5'-GTTTCTCTGCCGCGTTGT-3' | CMML-03 |
| *TET2* | 5'-TGCTGATGATGCTGATAATGC-3' | 5'-TCATTGTCCCTGCAGTCTGT-3' | CMML-03 |
| *VPS13D* | 5'-GGATGTAAAACCAATGCAGC-3' | 5'-CCCAAGACAGGATCTAAGGTG-3' | CMML-03 |
| *ZRSR2* | 5'-CCACCATGCCTGGTCTAAAG-3' | 5'-TGTGTCCCAGCTCTCTTGTG-3' | CMML-03 |
| *C2orf67* | 5'-TTGTCAAGGCATTGCTTAGTTC-3' | 5'-GCCTTGCCTGCAATCAG-3' | CMML-04 |
| *HTR2A* | 5'-CTAATTTGGCCCGTGTGC-3' | 5'-TGATGATGCTCTTGCCTTCA-3' | CMML-04 |
| *KRT39* | 5'-TGTCAGGTGCAACTCAAAGG-3' | 5'-GCTTGGCTTGGTTTGCAT-3' | CMML-04 |
| *LCA5* | 5'-TTAAGCGTTCTTTGAGTGCTG-3' | 5'-ATTTCGCTCCCAGAGCCT-3' | CMML-04 |
| *SFRS2* | 5'-AGACGCCATTTCCCCAG-3' | 5'-TTCGAGAAGTACGGGCG-3' | CMML-04 |
| *SKIV2L2* | 5'-CAATGGAACATGTTGAATCCAG-3' | 5'-TATGGCCCTGGCAAAATC-3' | CMML-04 |
| *STEAP2* | 5'-ATGGGAAGCCCTAAGAGCC-3' | 5'-CAGGTCCCACAGGGAGGTAT-3' | CMML-04 |
| *SYT10* | 5'-GGAAGAAACAGAAACAAGAACCC-3' | 5'-GCTAACATGCCGATTGCC-3' | CMML-04 |
| *SYTL2* | 5'-TCCTGTAAGATTGACAGTTGGAGA-3' | 5'-TCCATGATCCTTGTGCTAGGT-3' | CMML-04 |

**Continued from previous page**

| | | | |
|---|---|---|---|
| *TCN1* | 5'-TGCAAATGCATTCTTACTGC-3' | 5'-TTGCTCTGGGCTTTTCTG-3' | CMML-04 |
| *TMEM87B* | 5'-TGGAAGACTTACTGGTTGGAAA-3' | 5'-GCATTTGTTTTGCTTATCTCCC-3' | CMML-04 |
| *UGCG* | 5'-CCTTGTCTTTAGTTTGACAGTGAGTG-3' | 5'-ATGCAAGTGCCATGCAAA-3' | CMML-04 |

**Supplementary Table 9. List of PCR primers used for pooled target sequencing and validation**

| Target | Forward Primer | Reverse Primer |
|---|---|---|
| PRPF40B_ex1 | AAGCGGCCGCCAGGAAATCGGGCTTGAA | AAGCGGCCGCAAGGTCAGCAAAAGGGGC |
| PRPF40B_ex2 | AAGCGGCCGCGCCTTGGAATTGCCTCAG | AAGCGGCCGCTGCCCGCTCTTATCTCCA |
| PRPF40B_ex3 | AAGCGGCCGCTGCCCAAGTGAGGGTCAT | AAGCGGCCGCTGAGTACATGGCTGGCGA |
| PRPF40B_ex4 | AAGCGGCCGCTGAAGGAATGAGCCTCGG | AAGCGGCCGCTCCCAACAGGCATGATCC |
| PRPF40B_ex5 | AAGCGGCCGCTCTTTCTTCGGCAGACGG | AAGCGGCCGCCCCCTTTTCCAAAGCCTC |
| PRPF40B_ex6 | AAGCGGCCGCAGATGGAGGTCGGAGCAA | AAGCGGCCGCCAGGGCAGCTGAAAGGAA |
| PRPF40B_ex7 | AAGCGGCCGCCCACCTTCCCCAGTTTGA | AAGCGGCCGCTCACAGGGGCAATAGGGA |
| PRPF40B_ex8 | AAGCGGCCGCGCCCAAGGATCTGGATGA | AAGCGGCCGCTGCTGCTGCTGTTTCCTG |
| PRPF40B_ex9 | AAGCGGCCGCCCTTGTGGAGTCATCCCCT | AAGCGGCCGCCTGAGGCCAACATTGGGA |
| PRPF40B_ex10 | AAGCGGCCGCGGGGGTGATATGTTGCCA | AAGCGGCCGCTTCATGGCCTGTTCCCAT |
| PRPF40B_ex11 | AAGCGGCCGCACAGCATCAGCCACAGCA | AAGCGGCCGCGCACCCCAACCCAATCTA |
| PRPF40B_ex12 | AAGCGGCCGCGCAGCCTGCAGTGTCTCAT | AAGCGGCCGCTCTCAGGGTTCCCAAGTCC |
| PRPF40B_ex13 | AAGCGGCCGCAATGCCTACAAGGCGCAG | AAGCGGCCGCGCTGAAGGATTCTGCCCA |
| PRPF40B_ex14 | AAGCGGCCGCTTTCCCCAGGTCCTCCTCT | AAGCGGCCGCATCCCCGGAGTCCAAAGA |
| PRPF40B_ex15 | AAGCGGCCGCACCATTGCAGTGGCTGCT | AAGCGGCCGCGCAGGGAATGGGTGAAGA |
| PRPF40B_ex16 | AAGCGGCCGCTGCCTCTTCCTCTGCCAT | AAGCGGCCGCAGAGGGGTGGAGCCTGTAA |
| PRPF40B_ex17 | AAGCGGCCGCACTGATGTCCGCTTTGCC | AAGCGGCCGCAGTGTGCCCTGCATCCAT |
| PRPF40B_ex18 | AAGCGGCCGCCAAGGGCCTTCTTGACCA | AAGCGGCCGCCACCATCCCCTGAATTGG |
| PRPF40B_ex19 | AAGCGGCCGCCAATTCAGGGGATGGTGG | AAGCGGCCGCCTGGGGCAGAGAAAGGAAG |
| PRPF40B_ex20 | AAGCGGCCGCCTTCCTTTCTCTGCCCCAG | AAGCGGCCGCAATCTGGCTTTGCCTCCC |
| PRPF40B_ex21 | AAGCGGCCGCGGGAGGCAAAGCCAGATT | AAGCGGCCGCAAGCCCTCGTCTGGGAAT |
| PRPF40B_ex22 | AAGCGGCCGCAGGATTCCTTTGGCCCTG | AAGCGGCCGCAAAGGAGCTGGGGCTTGT |
| PRPF40B_ex23 | AAGCGGCCGCCCCAAGCAAGAAGCTGGA | AAGCGGCCGCCATCGCTCTCCTTGCCA |
| PRPF40B_ex24 | AAGCGGCCGCCGGAAAGCCAAGAAACCA | AAGCGGCCGCTGTTGGAGCAAAGGCACA |
| PRPF40B_ex25 | AAGCGGCCGCTTCCCCAGGCTTTGGAAT | AAGCGGCCGCTGGGGGTGGGGTTACTTT |
| SF1_ex1 | AAGCGGCCGCTGGAGGCAAAGATGAGGC | AAGCGGCCGCCCTGGCCAGTTCCAATGT |
| SF1_ex2 | AAGCGGCCGCGCAGATCTTGCTCAGAATAAATCG | AAGCGGCCGCGCTTCACAGTCTGCTGAGTTTCT |
| SF1_ex3 | AAGCGGCCGCTGTTGCCACCAGTAAAACG | AAGCGGCCGCTGGGGGTTTTTGCTTCAA |
| SF1_ex4 | AAGCGGCCGCCAGGTCACCCATGATCTGC | AAGCGGCCGCTCCGTGTGGGAATGAGAGA |
| SF1_ex5 | AAGCGGCCGCCCAATTAGGCACAAAACGG | AAGCGGCCGCTCCAGAAATGTGCGGTGA |
| SF1_ex6 | AAGCGGCCGCCCTAGGTTTCTGCTAGGAGGC | AAGCGGCCGCTCCAGCCTCTCTTCCCTCT |
| SF1_ex7 | AAGCGGCCGCGGTCAATGCAAGGGAACAA | AAGCGGCCGCTGAGCCCAGAAAGGAAGC |
| SF1_ex8 | AAGCGGCCGCGTTGTCCTGCACATCGCA | AAGCGGCCGCTTGGCTCGCTTAAATGGG |
| SF1_ex9 | AAGCGGCCGCATGGGGAAGCCAGACCA | AAGCGGCCGCTTGGTTCCTCAGGATAAAAGC |
| SF1_ex10 | AAGCGGCCGCTATCCCAAAGCCCCCAAG | AAGCGGCCGCCCGAAGAACAATGTTGCCA |
| SF1_ex11 | AAGCGGCCGCTGGAGGCAAAGATGAGGC | AAGCGGCCGCCCTGGCCAGTTCCAATGT |
| SF1_ex12 | AAGCGGCCGCCCCCACAAGCCCAAAATA | AAGCGGCCGCCATCTTTGCCTCCACCTGTC |
| SF1_ex13-1 | AAGCGGCCGCGCCCATCATGGTGACAAAGT | AAGCGGCCGCTACGACGACTACCACCACGA |
| SF1_ex13-2 | AAGCGGCCGCATAGTGGGGTTGCCTTGCAT | AAGCGGCCGCGTGACCTTGGGCTTCTTTACC |
| SF1_ex14 | AAGCGGCCGCACGATATGGAGCCAGCGT | AAGCGGCCGCTCACCCAACTCCCTTTGC |
| SF3A1_ex1 | AAGCGGCCGCGAGAGGCAAAATGGCAGG | AAGCGGCCGCCAGAGCGCAAACAGACCA |
| SF3A1_ex2 | AAGCGGCCGCCCCAAGCCAACAACGTCT | AAGCGGCCGCAACCTGCTCTGAACCAGGG |
| SF3A1_ex3 | AAGCGGCCGCCTCAAGCCCTGCGTAAGTG | AAGCGGCCGCAGGCACAAGGGAGTGGTGT |
| SF3A1_ex4 | AAGCGGCCGCTGTTTTCCTTTCCTGGGC | AAGCGGCCGCTGTTCACCCCTGTGCCT |
| SF3A1_ex5 | AAGCGGCCGCCATAAAACAGAACCAACCACCA | AAGCGGCCGCGCTGCCTCTTTGTACTGCG |
| SF3A1_ex6 | AAGCGGCCGCTGCTCTCTGGGAACACCTG | AAGCGGCCGCTGCTTCAGCGTAGCACCA |
| SF3A1_ex7 | AAGCGGCCGCGGCTTCAGGGGAAAAGGT | AAGCGGCCGCCAGGGCAAATAGGTGGGTC |
| SF3A1_ex8 | AAGCGGCCGCACAGGACTTCCTGGGCA | AAGCGGCCGCTCTGAGTCCTGCCATTGCT |
| SF3A1_ex9 | AAGCGGCCGCTCGGAGACTACCATGGGCT | AAGCGGCCGCCGCATGAAATACCACCCTCT |
| SF3A1_ex10 | AAGCGGCCGCCAAGTATGGAAGACGGGGG | AAGCGGCCGCATCCCATGGAACCAGCCT |
| SF3A1_ex11 | AAGCGGCCGCTCAGCCAACAGTGTCCGA | AAGCGGCCGCCCCCGTCTTCCATACTTGG |
| SF3A1_ex12 | AAGCGGCCGCCCCTGAACTAATGGCCTGG | AAGCGGCCGCAGTCCAACTTTGGCTTCCTG |
| SF3A1_ex13 | AAGCGGCCGCTCGCAAGTGTACGACAGGG | AAGCGGCCGCTCTCATCAGCCCTGAGTCCT |
| SF3A1_ex14 | AAGCGGCCGCTTTCCCCTGAATGCTTGG | AAGCGGCCGCCAGTGGTGAATCAGGCCA |
| SF3A1_ex15 | AAGCGGCCGCTGGATTCAGATGGCCACA | AAGCGGCCGCCCAAGCATTCAGGGGAAA |
| SF3A1_ex16 | AAGCGGCCGCGAGAGGCAAAATGGCAGG | AAGCGGCCGCCAGAGCGCAAACAGACCA |
| SF3B1_ex1 | AAGCGGCCGCGCTAGCAACGAAGAAAACACG | AAGCGGCCGCAGCCCCCAGCTATTTTTCTC |
| SF3B1_ex2 | AAGCGGCCGCCAGATGGCTGCAACAAAAGT | AAGCGGCCGCGGGAGTTCTTTTGGAAAGTGA |
| SF3B1_ex3 | AAGCGGCCGCTGGGAACTCAGACATTCACTTT | AAGCGGCCGCTGAAGGAGGGCTTAGACATCA |
| SF3B1_ex4 | AAGCGGCCGCTGGCATCTGTGACTTCAACC | AAGCGGCCGCCCTTGTTTCCATGTGCCTTA |
| SF3B1_ex5 | AAGCGGCCGCTGGGGGTAAGATTCTTTCTCAG | AAGCGGCCGCGGGCAGATAAATCAGTTGAACC |
| SF3B1_ex6 | AAGCGGCCGCGGCAACCCAAGCAGACTAAT | AAGCGGCCGCGAACCAGCGAGCGTAGAAAG |
| SF3B1_ex7 | AAGCGGCCGCACGTGTCCACCCAGGAAT | AAGCGGCCGCTTTCTGTGTGGGTGTGTGAAA |
| SF3B1_ex8 | AAGCGGCCGCCACAGAATACCACAAAGGAAAAA | AAGCGGCCGCCCCACCACACCTGGCTACTTT |
| SF3B1_ex9 | AAGCGGCCGCAGCCAATGCAAGCTTAATCA | AAGCGGCCGCCATGGCTACCCCTACTCCAG |
| SF3B1_ex10 | AAGCGGCCGCTCTCTCTTTTTGCTCTTCTGGAC | AAGCGGCCGCCAAGATAGATGTTGTCTCCAAGCA |
| SF3B1_ex11 | AAGCGGCCGCACATGGCAAAACCCTGTCTC | AAGCGGCCGCAATGACCAGCCATCTGGAAA |
| SF3B1_ex12 | AAGCGGCCGCTGCAAAGGAAAAGGTCTAGGA | AAGCGGCCGCTGGATGGAAATGAACTCATGC |
| SF3B1_ex13 | AAGCGGCCGCGTAGCCAGACCAGCAGCCTA | AAGCGGCCGCTCTTAAACAGTTCGTCCCTTGA |

**Continued from previous page**

| | | |
|---|---|---|
| SF3B1_ex14 | AAGCGGCCGCGGGCAACATAGTAAGACCCTGT | AAGCGGCCGCTAGAGTGGAAGGCCGAGAGA |
| SF3B1_ex15 | AAGCGGCCGCAGCAGCCAAACCCTATTTTT | AAGCGGCCGCAGGTAATGTTGGGGCATAGTT |
| SF3B1_ex16 | AAGCGGCCGCCCATCTGTCCCACAACACTG | AAGCGGCCGCCAGAGGAAAGGTAAATCCACCA |
| SF3B1_ex17 | AAGCGGCCGCGGCTGACTAAAATCCATCTCC | AAGCGGCCGCTCTCTTCATTTCAGGTCAGTTG |
| SF3B1_ex18 | AAGCGGCCGCTGGTAACCCCCTGAGCAT | AAGCGGCCGCCCTTGGAAAAGCAGTCTAAAAGG |
| SF3B1_ex19+20 | AAGCGGCCGCGTGACATCCCCCGACTTAAC | AAGCGGCCGCTCAACTGTGCAGTCATAAACCA |
| SF3B1_ex21 | AAGCGGCCGCACCCCCATTACCATCCAAAT | AAGCGGCCGCATCTGGGGCTTTCTCTTTCC |
| SF3B1_ex22 | AAGCGGCCGCTGCATTATTCAGACCATGCC | AAGCGGCCGCAAGACTAACTCTTCAGCCTTTCTGA |
| SF3B1_ex23 | AAGCGGCCGCCCTAAGACTCCAGGCTAGCAA | AAGCGGCCGCCAACAATGCCAGTGACTTGA |
| SF3B1_ex24 | AAGCGGCCGCTGCACGTTCAGCACAAGTATC | AAGCGGCCGCAAGGATTTTTCCCCCTTTCT |
| SF3B1_ex25 | AAGCGGCCGCAGGTGTGAAGTAGCTGTGCATT | AAGCGGCCGCCAACTCCCCAAGGGAAATAA |
| SRSF1_ex1 | AAGCGGCCGCTTGGAGCCTTCCCCAACT | AAGCGGCCGCTAAAGGCGCCATTTTGGA |
| SRSF1_ex2 | AAGCGGCCGCTCCCTTCACATCAATCCACA | AAGCGGCCGCCCCTTTTACTTTCTGCGATCC |
| SRSF1_ex3 | AAGCGGCCGCTGGACAACCTTGCCTGAA | AAGCGGCCGCCCTTACCAGCTTAACGTGAAAT |
| SRSF1_ex4 | AAGCGGCCGCGCAATTCAACACTTTAGCCCA | AAGCGGCCGCGCATGTCTGAAGATAGATGAAAGC |
| SRSF2_ex1-1 | AAGCGGCCGCAGGGGGAAAATGCAGACG | AAGCGGCCGCATCTCGGTCGACCTCCAA |
| SRSF2_ex1-2 | AAGCGGCCGCAGACGCCATTTCCCCAGT | AAGCGGCCGCCTCCCGATGTGGAGGGTAT |
| SRSF2_ex2 | AAGCGGCCGCACCCCAAATCCCATTTATGA | AAGCGGCCGCTTAACCCCGCTGTGCTTG |
| U2AF35_ex1 | AAGCGGCCGCCCCGAAAAGAGACCCCAT | AAGCGGCCGCGGTTCCCAGGACCGAAAA |
| U2AF35_ex2 | AAGCGGCCGCAGTCGATCACCTGCCTCACT | AAGCGGCCGCGGAGGTGCTTAATACCACGG |
| U2AF35_ex3 | AAGCGGCCGCTCAGCAGCTCAGAAGCCA | AAGCGGCCGCTGCACAGCAGCCAGTAGAA |
| U2AF35_ex4 | AAGCGGCCGCAGTGAAACGCGGTCCTTG | AAGCGGCCGCGGCAGCCGGTTAAGTGTTT |
| U2AF35_ex5 | AAGCGGCCGCCAAACAGATACAGAAGATCAACAGG | AAGCGGCCGCCAGCTCTGCCAGGCTTTT |
| U2AF35_ex6 | AAGCGGCCGCTGTCTCGCTTTCGCCTGT | AAGCGGCCGCGATCTTCTGTATCTGTTTGGAAGGA |
| U2AF35_ex7 | AAGCGGCCGCGAGACATTTACTACCTCGTGTGC | AAGCGGCCGCTTAAAGCGTGGATGGCAAG |
| U2AF35_ex8 | AAGCGGCCGCGGGGAGAAAGAAGCCACTG | AAGCGGCCGCTGAGATGGGGTGAGTGAGG |
| U2AF35_ex9 | AAGCGGCCGCCCTCCAAGATGACTAGCAGATTT | AAGCGGCCGCTGTGGTTCTGTGCTGCGT |
| U2AF65_ex1 | AAGCGGCCGCGAGGCGAAAGCTGCACA | AAGCGGCCGCGGGAAGCGAGGGAGAAAA |
| U2AF65_ex2 | AAGCGGCCGCCCTTACTCCGCTTATCCCG | AAGCGGCCGCGAACGAACGAGCCACACA |
| U2AF65_ex3 | AAGCGGCCGCTGTGGCATGTGGGGAAG | AAGCGGCCGCGGAAGGTGAAGGGAGTGGA |
| U2AF65_ex4 | AAGCGGCCGCGTTGGTCAGACTGAGGTTGC | AAGCGGCCGCAGAATCCCAACCCCTGTACC |
| U2AF65_ex5 | AAGCGGCCGCGGGGTTTTATCCGGCTTTTA | AAGCGGCCGCGGGTAGAGGGACAAGGGAGA |
| U2AF65_ex6 | AAGCGGCCGCACCCCTCTGTGTCTCCCTCT | AAGCGGCCGCACACGTTTCCTGGACCTCAC |
| U2AF65_ex7 | AAGCGGCCGCGGAAATCCCAATCCTGGAA | AAGCGGCCGCAGGCCATGGGAACAGGA |
| U2AF65_ex8 | AAGCGGCCGCTGGGAGATAACCTGGTACTGG | AAGCGGCCGCTCCATCAGGGCTGGGAT |
| U2AF65_ex9 | AAGCGGCCGCATCCCAGCCCTGATGGA | AAGCGGCCGCAGACCATGGCTTGGCAAC |
| U2AF65_ex10 | AAGCGGCCGCCATGGCGTTGGCTTTTTC | AAGCGGCCGCTCCCTCAAGCCAATGGAA |
| U2AF65_ex11 | AAGCGGCCGCCTTGAGCAGAGGGAGACTGG | AAGCGGCCGCAGGGTGAGAGAGCACTGAGC |
| U2AF65_ex12 | AAGCGGCCGCGGCGTGAGCCTTGTTCA | AAGCGGCCGCTCCTCTGCCCATCGTCTT |
| ZRSR2_ex1 | AAGCGGCCGCCGTTTCAAGTCCCACGCT | AAGCGGCCGCTTCCTGCCACCACATCCT |
| ZRSR2_ex2 | AAGCGGCCGCCTGCATTGTAGCCGCTGA | AAGCGGCCGCGCTGGAGTGGACAGAGCAA |
| ZRSR2_ex3 | AAGCGGCCGCTTGACCAAGGATTTGCAGC | AAGCGGCCGCGACTGGTACTGGTTAGTAAAGGTTGA |
| ZRSR2_ex4 | AAGCGGCCGCTGTGGATTAATGCCCATTTC | AAGCGGCCGCCCAACCTCCCAAGATAGGC |
| ZRSR2_ex5 | AAGCGGCCGCTGTGCGCTGTATGTGAAATG | AAGCGGCCGCCCCAAACTCTGACATGCCT |
| ZRSR2_ex6 | AAGCGGCCGCTGTGTGCGTGTGTGTGTTTT | AAGCGGCCGCCCACGAAACTAACATTACTGGAAC |
| ZRSR2_ex7 | AAGCGGCCGCTCATGGGTTTTTACTCCACCA | AAGCGGCCGCCCTCTCCCAAAAGGGGAA |
| ZRSR2_ex8 | AAGCGGCCGCCCACCATGCCTGGTCTAAAG | AAGCGGCCGCTGTGTCCCAGCTCTCTTGTG |
| ZRSR2_ex9 | AAGCGGCCGCGGGGAATGTTAGCCTGGA | AAGCGGCCGCCAGGAAGACATCCACAAGCA |
| ZRSR2_ex10 | AAGCGGCCGCCAGTGAACTTGGTGGTCCTACA | AAGCGGCCGCACTGGGTTTCCCCCAAAG |
| ZRSR2_ex11-1 | AAGCGGCCGCCAGAATATGCAGTGATAAGTGCTG | AAGCGGCCGCAACTTTGGCTCCGGCTG |
| ZRSR2_ex11-2 | AAGCGGCCGCGGCCACCACGACGACTACTA | AAGCGGCCGCGCTCCAATGTTTAGTTCTCTCTCAG |